重点大学信息安全专业规划系列教材

计算机网络安全

彭飞　龙敏　编著

清华大学出版社

重点大学信息安全专业规划系列教材

计算机网络安全

彭飞 龙敏 编著

清华大学出版社
北京

内 容 简 介

本书全面地介绍了计算机网络安全技术。全书共分为12章，第1章介绍计算机网络安全的基本概念、内容和方法，随后的11章分别从网络协议安全、信息加密与认证、访问控制、防火墙与入侵检测、数据备份与恢复、操作系统安全、Web站点安全、电子邮件安全、无线网络安全、恶意软件攻击与防治以及网络入侵与取证等不同层面对计算机网络安全的相关理论与方法进行了详细介绍。

本书适合作为信息安全专业本科高年级学生以及研究生的专业课教材，也可供从事信息安全专业的技术人员阅读参考。

图书在版编目(CIP)数据

计算机网络安全/彭飞，龙敏编著. —北京：清华大学出版社，2013.4
重点大学信息安全专业规划系列教材
ISBN 978-7-302-31125-6

Ⅰ. ①计… Ⅱ. ①彭… ②龙… Ⅲ. ①计算机网络－安全技术－高等学校－教材 Ⅳ. ①TP393.08

中国版本图书馆CIP数据核字(2012)第315242号

责任编辑： 魏江江 王冰飞
封面设计：
责任校对： 白 蕾
责任印制：

出版发行： 清华大学出版社
网　　址： http://www.tup.com.cn，http://www.wqbook.com
地　　址： 北京清华大学学研大厦A座　　**邮　　编：** 100084
社 总 机： 010-62770175　　**邮　　购：** 010-62786544
投稿与读者服务： 010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈： 010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载： http://www.tup.com.cn，010-62795954
印 刷 者：
装 订 者：
经　　销： 全国新华书店
开　　本： 185mm×260mm　　**印　　张：** 19.75　　**字　　数：** 483千字
版　　次： 2013年4月第1版　　**印　　次：** 2013年4月第1次印刷
印　　数： 1～ 000
定　　价： .00元

产品编号：045677-01

出版说明

由于网络应用越来越普及,信息化的社会已经呈现出越来越广阔的前景,可以肯定地说,在未来的社会中电子支付、电子银行、电子政务以及多方面的网络信息服务将深入到人类生活的方方面面。同时,随之面临的信息安全问题也日益突出,非法访问、信息窃取、甚至信息犯罪等恶意行为导致信息的严重不安全。信息安全问题已由原来的军事国防领域扩展到了整个社会,因此社会各界对信息安全人才有强烈的需求。

信息安全本科专业是2000年以来结合我国特色开设的新的本科专业,是计算机、通信、数学等领域的交叉学科,主要研究确保信息安全的科学和技术。自专业创办以来,各个高校在课程设置和教材研究上一直处于探索阶段。但各高校由于本身专业设置上来自于不同的学科,如计算机、通信和数学等,在课程设置上也没有统一的指导规范,在课程内容、深浅程度和课程衔接上,存在模糊不清、内容重叠、知识覆盖不全面等现象。因此,根据信息安全类专业知识体系所覆盖的知识点,系统地研究目前信息安全专业教学所涉及的核心技术的原理、实践及其应用,合理规划信息安全专业的核心课程,在此基础上提出适合我国信息安全专业教学和人才培养的核心课程的内容框架和知识体系,并在此基础上设计新的教学模式和教学方法,对进一步提高国内信息安全专业的教学水平和质量具有重要的意义。

为了进一步提高国内信息安全专业课程的教学水平和质量,培养适应社会经济发展需要的、兼具研究能力和工程能力的高质量专业技术人才。在教育部相关教学指导委员会专家的指导和建议下,清华大学出版社与国内多所重点大学共同对我国信息安全人才培养的课程框架和知识体系,以及实践教学内容进行了深入的研究,并在该基础上形成了"信息安全人才需求与专业知识体系、课程体系的研究"等研究报告。

本系列教材是在课程体系的研究基础上总结、完善而成,力求充分体现科学性、先进性、工程性,突出专业核心课程的教材,兼顾具有专业教学特点的相关基础课程教材,探索具有发展潜力的选修课程教材,满足高校多层次教学的需要。

本系列教材在规划过程中体现了如下一些基本组织原则和特点。

(1) 反映信息安全学科的发展和专业教育的改革,适应社会对信息安全人才的培养需求,教材内容坚持基本理论的扎实和清晰,反映基本理论和原理的综合应用,在其基础上强调工程实践环节,并及时反映教学体系的调整和教学内容的更新。

(2) 反映教学需要,促进教学发展。教材要适应多样化的教学需要,正确把握教学内容和课程体系的改革方向,在选择教材内容和编写体系时注意体现素质教育、创新能

力与实践能力的培养，为学生知识、能力、素质协调发展创造条件。

(3) 实施精品战略，突出重点。规划教材建设把重点放在专业核心(基础)课程的教材建设上；特别注意选择并安排一部分原来基础比较好的优秀教材或讲义修订再版，逐步形成精品教材；提倡并鼓励编写体现工程型和应用型的专业教学内容和课程体系改革成果的教材。

(4) 支持一纲多本，合理配套。专业核心课和相关基础课的教材要配套，同一门课程可以有多本具有各自内容特点的教材。处理好教材统一性与多样化，基本教材与辅助教材、教学参考书，文字教材与软件教材的关系，实现教材系列资源的配套。

(5) 依靠专家，择优落实。在制定教材规划时依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时，要引入竞争机制，通过申报、评审确定主编。书稿完成后认真实行审稿程序，确保出书质量。

繁荣教材出版事业，提高教材质量的关键是教师。建立一支高水平的、以老带新的教材编写队伍才能保证教材的编写质量，希望有志于教材建设的教师能够加入到我们的编写队伍中来。

21 世纪高等学校信息安全专业规划教材

联系人：魏江江 weijj@tup.tsinghua.edu.cn

前　言

随着信息技术与 Internet 的发展，计算机网络在给人们的生活和工作带来便利的同时，也面临着严重的安全威胁，例如非法侵入计算机系统窃取机密信息、篡改和破坏数据、病毒、蠕虫、垃圾邮件、僵尸网络等。网络安全已关系到国家安全和社会稳定等重要问题。

计算机网络安全作为信息安全领域的一个重要方面，其相关技术还在不断地研究与发展。本书作者结合所在单位信息安全专业本科生和相关方向研究生培养的实际情况，编撰和出版本书作为专业课程教材。

全书共分为 12 章，第 1 章介绍了计算机网络安全的基本概念、内容和方法，分析网络安全问题产生的根源，并对安全问题进行分类；另外，还介绍了网络安全的等级标准。第 2 章介绍了网络层安全协议 IPSec，传输层安全协议 SSL、TLS 和应用层安全协议 SET、Telnet、HTTP 等。第 3 章介绍了密码学的基本原理，主要包括古典密码技术、对称密码技术以及非对称密码技术、信息认证的基本概念，单向 Hash 函数与消息认证码的基本原理及典型的认证方法和技术。第 4 章介绍了访问控制的概念、模型及访问控制中涉及的 AAA 技术与 VPN 技术。第 5 章介绍了防火墙的基本概念和种类、防火墙的体系结构及功能、入侵检测技术的种类及各类技术的相关性能。第 6 章介绍了数据的基本概念及数据备份与恢复所需的相关基础知识、数据备份技术与数据恢复技术。第 7 章介绍了计算机操作系统安全的基本概念，包括安全操作系统评价标准、常见的系统安全保护方法、单点登录的访问管理以及主流操作系统的主要安全机制等方面的知识。第 8 章介绍了 Web 应用程序上的安全问题及漏洞，以及基于 IIS 和 ASP 网站安全体系的建立及技术实现。第 9 章介绍了电子邮件安全的基本特性及其面临的安全问题，对电子邮件的几种安全技术进行了全面介绍，包括 PGP、S/MIME、PEM、PKI 技术及安全防范措施等。第 10 章分析了当前无线网络所面临的安全威胁及其防范措施，对于无线网络的代表性技术 IEEE 802.11 安全和蓝牙安全进行了介绍。第 11 章对恶意软件的基本知识、恶意软件的相关危害及防治方法和典型恶意软件的攻防方法进行了介绍。第 12 章分析了导致网络脆弱的因素，对网络入侵的常用方法及防范措施、入侵检测系统的原理、结构和流程以及计算机取证的一般步骤和取证模型进行了介绍。

本书由彭飞负责编写，全书由龙敏负责整理修订。本书适合作为信息安全专业本科高年级学生以及研究生的专业课教材，也可供从事信息安全专业的技术人员和研究人员阅读参考。

本书作为教材适合于 48～64 学时的教学，建议的教学方式为课堂讲授与实验相结

合，教师可根据书上的思考题，指导学生进行编程或仿真实验，通过对原理和应用算法的实验，进一步加深学生对所学内容的理解。

本书作者多年来一直从事信息安全的教学和研究工作，本书也是网络与信息安全湖南省重点实验室全体师生多年从事数字内容安全研究工作成果的结晶。

在本书的编写过程中，钱勤、刘艳、陈丽、朱小文、李洪淋、刘娟、李姣婷等研究生参与了部分资料的收集与整理工作；清华大学出版社的魏江江主任为本书的高质量出版倾注了大量心血；此外，本书的编写还得到了湖南大学信息科学与工程学院李仁发教授、赵欢教授等的大力支持，在此对他们付出的辛勤劳动表示由衷感谢。

计算机网络安全技术日新月异，限于作者水平和经验，书中难免出现疏漏之处，望读者提出宝贵意见，以便再版时修改和完善。

编　者

2012年10月

目　　录

第 1 章　计算机网络安全概述

本章学习目标

计算机网络安全问题是随着信息技术和网络技术的发展而出现的，网络安全涉及各行各业的许多重大利益问题，因此计算机网络的安全防护已得到广泛重视。本章介绍计算机网络安全的基本概念、内容和方法，分析网络安全问题产生的根源，并对网络安全问题进行分类，介绍网络安全的等级标准。

通过对本章的学习，应掌握以下内容：

(1) 网络安全问题的产生与分类。

(2) 计算机网络安全的基本概念、内容、目标和要求。

(3) 计算机网络安全体系结构与基本方法。

(4) 计算机网络安全评估的概念与方法。

(5) 网络安全等级标准。

信息技术的发展给人们的生活、工作等方面带来了便捷和好处。然而，计算机信息技术是一把双刃剑，它在为人们提高工作效率，为社会创造更多财富的同时，也为一些人利用它进行非法勾当提供了可能。例如非法侵入计算机系统窃取机密信息、篡改和破坏数据等。这些非法行为将给社会造成难以估量的损失。据统计，全球约每 20 秒钟就有一次计算机入侵事件发生，Internet 上的网络防火墙约有 1/4 曾被突破过，大约有 70%以上的网络管理人员报告曾因机密信息泄露而受到了损失。

在当前的数字化时代，信息技术和网络技术与人们的生活和工作息息相关、密不可分。因此，网络安全已关系到国家安全和主权、社会的稳定、民族文化的继承和发扬等重要问题。网络安全的涉及面很广，包含了计算机科学、网络技术、通信技术、密码技术、信息安全技术、应用数学、数论、信息论等多种学科。同时，除了技术上的问题，还有法律的问题、管理的问题等。

1.1　计算机网络安全的基本概念

1.1.1　网络安全的定义

计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备通过通信线路连接起来，在网络操作系统、网络管理软件及网络通信协议的管理和协调下，实现资源共享和信息传输的计算机系统。

从一般意义来看，安全是指没有危险和不出事故。对于计算机网络而言，其安全问题是指网络系统的硬件、软件及其系统中的数据受到保护，不遭到偶然的或者恶意的原因破坏、更改、泄露，系统连续、可靠、正常地运行，网络服务不中断。从广义上来说，凡是涉及网络上信息的保密性、完整性、可用性、真实性和可控性的相关技术和理论都是网络安全所要研究

的领域。

计算机网络的安全实际上包括两方面的内容：一是网络的系统安全，二是网络的信息安全。由于计算机网络最重要的资源是它向用户提供的服务及其所拥有的信息，因而计算机网络的安全性可以定义为：保障网络服务的可用性和网络信息的完整性。前者要求网络向所有用户有选择地随时提供各自应得到的网络服务，后者则要求网络保证信息资源的保密性、完整性、可用性和准确性。可见，建立安全的网络系统要解决的根本问题是如何在保证网络的连通性、可用性的同时对网络服务的种类、范围等进行适当程度的控制从而保障系统的可用性和信息的完整性不受影响。

由此可见，网络安全涉及的内容既有技术方面的问题，也有管理方面的问题，二者相互补充，缺一不可。技术方面主要侧重于防范外部非法用户的攻击，管理方面则侧重于内部人为因素的管理。如何更有效地保护重要的信息数据、提高计算机网络系统的安全性已经成为所有计算机网络应用必须考虑和必须解决的一个重要问题。

1.1.2 网络安全的基本特征

一个安全的计算机网络通常应具有以下几个特点。

1. 保密性

保密性是指网络信息不被泄露的特性。保密性是保证网络信息安全的一个非常重要的手段。保密性可以保证即使信息泄露，非授权用户在有限的时间内也无法识别真正的信息内容。常用到的保密措施主要包括：信息加密和物理保密（限制、隔离、隐蔽、控制），防辐射，防监听等。

2. 完整性

完整性是指网络信息未经授权不能进行改变的特性，即网络信息在存储和传输过程中不被删除、修改、伪造、乱序、重放和插入等操作改变，保持信息的原样。影响网络信息完整性的主要因素包括设备故障、误码、人为攻击以及计算机病毒等。

3. 可用性

可用性是指网络信息可被授权用户访问的特性，即网络信息服务在需要时能够保证授权用户使用。这里包含两个含义：当授权用户访问网络时不致被拒绝；授权用户访问网络时要进行身份识别与确认，并且对用户的访问权限加以规定的限制。

4. 可控性

可控性是指可被授权实体访问并按需求使用的特性，即当需要时应能存取所需的信息。可控性要求能对信息的传播及内容具有控制能力。

5. 可靠性

可靠性是网络系统安全最基本的要求，主要是指网络系统硬件和软件无故障运行的性能。提高可靠性的具体措施主要包括：提高设备质量，配备必要的冗余和备份，采取纠错、自愈和容错等措施，强化灾害恢复机制，合理分配负荷等。

6. 不可抵赖性

不可抵赖性也称作不可否认性，主要用于网络信息的交换过程，保证信息交换的参与者

都不可能否认或抵赖曾进行的操作,类似于在发文或收文过程中的签名和签收的过程。

1.2　计算机网络面临的安全威胁

由于早期 Internet 仅为一个完全非营利性的信息共享平台,在安全机制方面的考虑较少。然而,随着 Internet 的全球化与商业化,安全问题越来越突出,因此安全性已成为人们广泛关注的问题。

1.2.1　影响网络安全的因素

计算机网络的安全隐患多数是利用网络系统本身存在的安全弱点,而在网络的使用、管理过程中的不当行为可能会进一步加剧安全问题的严重性。影响网络安全的因素有很多,归纳起来主要包括 3 个方面:技术因素、管理因素和人为因素。

1. 技术因素

从技术因素来看,主要包括硬件系统的安全缺陷、软件系统的安全漏洞和系统安全配置不当造成的其他安全漏洞 3 种情况。

(1) 硬件系统的安全缺陷。由于理论或技术的局限性,必然会导致计算机及其硬件设备存在这样或那样的不足,进而在使用时可能产生各种各样的安全问题。

(2) 软件系统的安全漏洞。在软件设计时期,人们为了能够方便不断改进和完善所涉及的系统软件和应用软件,开设了"后门"以便更新和修改软件的内容,这种后门一旦被攻击者掌握将成为影响系统安全的漏洞。同时,在软件开发过程中,由于结构设计的缺陷或编写过程的不规范也会导致安全漏洞的产生。

(3) 系统安全配置不当造成的其他安全漏洞。通常在系统中都有一个默认配置,而默认配置的安全性通常较低。此外,在网络配置时出现错误,存在匿名 FTP、Telnet 的开放、密码文件缺乏适当的安全保护、命令的不合理使用等问题都会导致或多或少的安全漏洞。黑客就有可能利用这些漏洞攻击网络,影响网络的安全性。

2. 管理因素

管理因素主要是指网络管理方面的漏洞。通常来说,很多机构在设计内部网络时,主要关注来自外部的威胁,对来自内部的攻击考虑较少,导致内部网络缺乏审计跟踪机制,网络管理员没有足够重视系统的日志和其他信息。另外,管理人员的素质较差、管理措施的完善程度不够以及用户的安全意识淡薄等都会导致网络安全问题。

3. 人为因素

安全问题最终根源都是人的问题。前面提到的技术因素和管理因素均可以归结到人的问题。根据人的行为可以将网络安全问题分为人为的无意失误和人为的恶意攻击。

(1) 人为的无意失误。此类问题主要是由系统本身故障、操作失误或软件出错导致的。例如管理员安全配置不当造成的安全漏洞、网络用户安全意识不强带来的安全威胁等。

(2) 人为的恶意攻击。此类问题是指利用系统中的漏洞而进行的攻击行为或直接破坏物理设备和设施的攻击行为。例如病毒可以突破网络的安全防御入侵到网络主机上,可能

造成网络系统的瘫痪等安全问题。

1.2.2 网络攻击类型

计算机网络的主要功能是传输信息，信息传输主要面临的威胁包括如下四类：

(1) 截获：攻击者从网络上窃听他人的通信内容。

(2) 中断：攻击者有意中断他人在网络上的通信。

(3) 篡改：攻击者故意篡改在网络上传输的报文。

(4) 伪造：攻击者伪造信息在网络上传输。

当前网络安全的威胁主要体现在以下几个方面：

(1) 网络协议中的缺陷：例如 TCP/IP 协议的安全问题等。

(2) 窃取信息：例如通过物理搭线、监视信息流、接收辐射信号、会话劫持、冒名顶替等形式窃取通信信息。

(3) 非法访问：通过伪装、IP 欺骗、重放、破译密码等方法滥用或篡改网络信息。

(4) 恶意攻击：通过拒绝服务攻击、垃圾邮件、逻辑炸弹、木马工具等中断网络服务或破坏网络资源。

(5) 黑客行为：由于黑客的入侵或破坏，造成非法访问、拒绝服务、计算机病毒、网络钓鱼等。

(6) 计算机病毒：例如利用病毒破坏计算机功能或破坏数据，影响计算机使用或破坏网络。

(7) 电子间谍活动：例如信息流量分析、信息窃取等。

(8) 信息战：通过利用、破坏敌方和保护己方的信息系统而展开的一系列作战活动。

(9) 人为行为：例如使用不当、安全意识差等。

根据攻击者对网络中信息是否进行更改，网络攻击可分为被动攻击和主动攻击。

(1) 被动攻击：攻击者非法截获、窃取通信线路中的信息，使信息保密性遭到破坏，信息泄露却无法察觉，从而给用户带来巨大的损失。

(2) 主动攻击：攻击者通过网络线路将虚假信息或计算机病毒传入信息系统内部，破坏信息的真实性、完整性及系统服务的可用性，即通过中断、伪造、篡改和重排信息内容造成信息破坏，使系统无法正常运行。

1.2.3 网络安全威胁的发展趋势

信息技术的发展极大地改变了人们的生活，随之而来的网络安全形势愈加严峻。安全攻击手段开始由简单化向综合化演变，攻击形式向多样化、复杂化发展。病毒、蠕虫、垃圾邮件、僵尸网络等攻击持续增长，各种软硬件安全漏洞被利用并进行攻击的综合成本越来越低，同时，内部人员的蓄意攻击也防不胜防。

网络安全威胁的发展趋势主要包括：

(1) 恶意软件的不断演变。木马、蠕虫、僵尸网络等恶意软件针对 Web 的攻击成为新的热点，这些恶意软件的攻击方式在未来也会有很多演进。

(2) P2P 应用引发新的安全问题。P2P 技术在给 Internet 带来极大促进的同时，也给网络应用带来了一些隐患，例如版权问题等。目前 P2P 流量高达整个主干网络流量的 40%

以上,这不仅造成了带宽的紧张,也影响了其他 Internet 业务的正常开展。如何正确优化带宽并合理使用 P2P 技术将成为未来人们的重要挑战。

(3) 新兴无线终端攻击的安全。随着 3G、Wimax、LTE 等多种无线宽带技术的推广应用,无线终端用户数目已超过固网用户数目。智能手机、无线数据卡等各种形式的移动终端已开始成为黑客攻击的主要目标。针对无线终端的攻击包括:基于彩信应用的蠕虫;针对手机操作系统的病毒攻击;恶意广播的垃圾电话;垃圾短信、彩信;手机信息被窃取;针对无线业务的木马攻击;SIM 卡复制以及针对无线传输协议的黑客攻击等。

(4) 数据泄露的新形势。随着新的存储介质、电子邮件、博客、微博、社交网站等各种新型信息传播工具的应用,数据泄露攻击也出现了新的形式:U 盘、移动硬盘、红外、蓝牙等传输模式携带或外传重要的敏感信息,均可能导致重要数据的泄露;通过对电子设备(例如 PC)重构电磁信息,实时获取重要信息;通过植入木马盗取主机介质或者外设上的重要信息数据;通过截获在公网传播的 E-mail 信息或无线传播的数据信息,获取敏感信息。针对信息获取的数据泄露攻击方式已成为攻击者实施攻击的重要方式。

(5)安全攻击新方向。新的信息技术应用也将会产生一些新型的安全攻击方法。例如:针对虚拟化技术应用产生的安全问题;针对安全专用软硬件的攻击;针对网络设备无线设备等通信设备的攻击;各种规模的分布式 DDoS 攻击;形形色色的 Web 应用攻击等。

1.3　计算机网络安全模型与体系结构

1.3.1　网络安全模型

网络安全模型是对动态网络安全过程的抽象描述。为了对网络上存在的威胁有效地进行安全防范,需要建立合理的网络安全模型以指导网络安全工作的部署和管理。目前,在网络安全领域存在比较多的网络安全模型。这些模型都能较好地描述网络安全的部分特征,但又有各自的侧重点,在不同的领域都有其各自的应用。通过对网络安全模型的研究,能充分了解网络安全动态过程的构成因素,有助于网络安全策略体系的合理构建。常见的网络安全模型包括 PPDR 模型和 APPDRR 模型。

1. PPDR 模型

PPDR 模型是由美国国际 Internet 安全系统公司(Internet Security Systems Inc.,ISS)提出的一个可适应网络动态安全模型。PPDR 模型中包括 4 个非常重要的环节:策略(Policy)、防护(Protection)、检测(Detection)和响应(Response)。如图 1-1 所示,防护、检测和响应组成了一个完整的、动态的安全循环,在策略的指导下保证网络系统的安全。



图 1-1　PPDR 模型

2. APPDRR 模型

虽然网络安全的动态性在 PPDR 模型中得到了一定程度的体现,但该模型不能描述网络安全的动态螺旋上升过程。为此,人们对 PPDR 模型进行了改进,在此基础上提出了 APPDRR 模型。APPDRR 模型认为网络安全由评估(Assessment)、策略(Policy)、

防护(Protection)、检测(Detection)、响应(Reaction)和恢复(Restoration)6 个部分组成。

根据 APPDRR 模型,网络安全的第一个重要环节是对网络进行风险评估,掌握所面临的风险信息。策略是 APPDRR 模型的第二个重要环节,它起着承上启下的作用:一方面,安全策略应当随着风险评估的结果和安全需求的变化做相应的更新;另一方面,安全策略在整个网络安全工作中处于原则性的指导地位,其后的检测、响应诸环节都应在安全策略的基础上展开。防护是安全模型中的第三个环节,体现了网络安全的防护措施。接下来是动态检测、实时响应、灾难恢复 3 个环节,体现了安全动态防护与安全入侵、安全威胁进行对抗的特征。

APPDRR 模型将网络安全视为一个不断改进的过程,即通过风险评估、安全策略、系统防护、动态检测、实时响应和灾难恢复 6 个环节,使网络安全得以完善和提高。

1.3.2 ISO/OSI 安全体系结构

国际标准化组织(International Standard Organization,ISO)制定了开放系统互联(Open System Interconnection,OSI)网络参考模型。OSI 参考模型定义了开放系统的层次结构、层次之间的相互关系及各层所包含的可能的服务。它是作为一个框架来协调和组织各层协议的制定,也是对网络内部结构最精练的概括与描述。

ISO/OSI 安全体系是根据 OSI 七层协议模型建立的,即 OSI 安全体系结构与 OSI 七层是相对应的,在不同的层次上都有不同的安全技术。ISO/OSI 安全体系包括安全服务、安全机制、安全管理和安全层次 4 部分的内容。这 4 个部分是一个联系紧密的整体,其中,安全机制是核心,安全服务和安全管理通过安全机制实现,安全服务的位置由安全层次来描述。

1. 安全服务

ISO/OSI 安全体系结构定义了一组安全服务,主要包括认证服务、访问控制服务、数据保密服务、数据完整性服务和抗否认性服务。

2. 安全机制

ISO/OSI 安全体系结构分为八大类安全机制,分别包括加密机制、数据签名机制、访问控制机制、数据完整性机制、认证机制、业务流填充机制、路由控制机制和公正机制。

(1) 加密机制:是确保数据安全性的基本方法,在 ISO/OSI 安全体系结构中应根据加密所在的层次及加密对象的不同,而采用不同的加密方法。

(2) 数字签名机制:是确保数据真实性的基本方法,利用数字签名技术可进行用户的身份认证和消息(本书中消息与信息通用)认证,它具有解决收、发双方纠纷的能力。

(3) 访问控制机制:从计算机系统的处理能力方面对信息提供保护。访问控制按照事先确定的规则决定主体对客体的访问是否合法,当主体试图非法使用一个未经授权的客体(资源)时,访问控制机制将拒绝这一企图,给出报警并记录日志档案。

(4) 数据完整性机制:破坏数据完整性的主要因素有数据在信道中传输时受信道干扰影响产生错误、数据在传输和存储过程中被非法入侵者篡改、计算机病毒对程序和数据的传染等。纠错编码和差错控制是对付信道干扰的有效方法。对付非法入侵者主动攻击的有效方法是保密认证,对付计算机病毒有各种病毒检测、杀毒和免疫方法。

(5) 认证机制：在计算机网络中认证主要有用户认证、消息认证、站点认证和进程认证等，可用于认证的方法有已知信息（例如密码）、共享密钥、数字签名、生物特征（例如指纹）等。

(6) 业务流填充机制：攻击者通过分析网络中某一路径上的信息流量和流向来判断某些事件的发生，为了对付这种攻击，一些关键站点间在无正常信息传输时，持续传输一些随机数据，使攻击者不知道哪些数据是有用的，哪些数据是无用的，从而挫败攻击者的信息流分析。

(7) 路由控制机制：在大型计算机网络中，从源点到目的地往往存在多条路径，其中有些路径是安全的，有些路径是不安全的，路由控制机制可根据信息发送者的申请选择安全路径，以确保数据安全。

(8) 公正机制：在计算机网络中，并不是所有的用户都是诚实可信的，同时也可能由于设备故障等技术原因造成信息丢失、延迟等，用户之间很可能引起责任纠纷。为了解决这个问题，就需要有一个各方都信任的第三方提供公证仲裁，仲裁数字签名技术是这种公正机制的一种技术支持。

3. 安全管理

ISO/OSI安全体系结构的安全管理是实施一系列的安全政策，对系统和网络上的操作进行管理。它包括三部分内容：系统安全管理、安全服务管理和安全机制管理。

(1) 系统安全管理：涉及整体OSI安全环境的管理，包括总体安全策略的管理、OSI安全环境之间的安全信息交换、安全服务管理和安全机制管理的交互作用、安全事件的管理、安全审计管理和安全恢复管理。

(2) 安全服务管理：涉及特定安全服务的管理，包括对某种安全服务定义其安全目标、指定安全服务可使用的安全机制、通过适当的安全机制管理及调动需要的安全机制、系统安全管理以及安全机制管理相互作用。

(3) 安全机制管理：涉及特定的安全机制的管理，包括密钥管理、加密管理、数字签名管理、访问控制管理、数据完整性管理、鉴别管理、业务流填充管理等。

4. 安全层次

ISO/OSI安全体系是在不同的网络层上采用不同安全机制实现的，分布情况如表1-1所示。

表1-1　各网络安全层次的主要安全机制

网络安全层次	安全机制
应用层安全	身份认证、访问控制、数据加密、数字签名、交换认证、业务流分析
表示层安全	
会话层安全	
传输层安全	身份认证、访问控制、数据加密、路由控制、一致性检查
网络层安全	
数据链路层安全	数据加密、数据流加密
物理层安全	

1.4 网络安全等级

美国早在20世纪80年代就针对其国防部门的计算机安全保密开展了一系列有影响的工作，后来成立了国家计算机安全中心(National Computer Security Center，NCSC)继续进行有关工作。1983年他们公布了可信计算机系统评估准则(Trusted Computer System Evaluation Criteria，TCSEC)，其中使用了可信计算基础(Trusted Computing Base，TCB)这一概念，即计算机硬件与支持可信应用及可信用户的操作系统的组合体。在TCSEC的评价准则中，从B级开始就要求具有强制存取控制和形式化模型技术的应用。从网络安全的角度出发，TCSEC准则对用户登录、授权管理、访问控制、审计跟踪、隐通道分析、可信通道建立、安全检测、生命周期保障、文本写作、用户指南等均提出了规范性要求，并根据所采用的安全策略、系统所具备的安全功能将系统分为四类7个安全级别。将计算机系统的可信程度划分为D、C1、C2、B1、B2、B3和A1这7个层次，各层的特性如表1-2所示。

表1-2 TCSEC的四类7个安全级别

类别	级别	特　性
D	D	D类的安全级别最低，保护措施最少且没有安全功能
C	C1	自主安全保护级，它能够实现用户与数据的分离。数据的保护是以用户组为单位的，并实现对数据进行自主存取控制实现
	C2	受控访问级，该级可以通过登录规程、审计安全性相关事件等来隔离资源
B	B1	标记安全保护级。该级对系统的数据进行标记，同时对标记的主体和客体实行强制的存取控制
	B2	结构化安全保护级。该级建立形式化的安全策略模型，同时对系统内的所有主体和客体都实现强制访问和自主访问控制
	B3	安全级，它能够实现访问监控器的要求。访问监控器是指监控的主体和客体之间授权访问关系的部件。该级还支持安全管理员职能、扩充审计机制，当发生与安全相关的事件时将发出信号，同时可以提供系统恢复过程
A	A1	A1级的功能与B3级几乎是相同的，但是A1级的特点在于它的系统拥有正式的分析和数学方法，它可以完全证明一个系统的安全策略和安全规格的完整性与一致性。同时，A1级还规定了将完全计算机系统运送到现场安装所遵守的程序

TCSEC带动了国际计算机安全的评估研究，20世纪90年代西欧4国(英、法、荷、德)联合提出了信息技术安全评估标准(Information Technology Security Evaluation Criteria，ITSEC)。ITSEC除了吸收TCSEC的成功经验外，首次提出了信息安全的保密性、完整性、可用性的概念，把可信计算机的概念提高到可信信息技术的高度上来认识。他们的工作成为欧共体信息安全计划的基础，并对国际信息安全的研究、实施带来深刻的影响。

美国为了保持他们在制定准则方面的优势，不甘心TCSEC的影响被ITSEC取代，采取联合其他国家共同提出新评估准则的办法来体现其领导作用。1991年1月宣布的制定通用安全评估准则(Common Criteria，CC)的计划，其基础是欧洲的ITSEC，美国的包括TCSEC在内的新的联邦评估标准，加拿大的CTCPEC，以及国际标准化组织ISO：

SC27WG3 的安全评估标准。CC 标准吸收了各先进国家对现代信息系统信息安全的经验与知识，对其后及未来信息安全的研究与应用带来了重大影响。

我国公安部组织制订了《计算机信息系统安全保护等级划分准则》国家标准，并于 1999 年 9 月 13 日由国家质量技术监督局审查通过并正式批准发布，已于 2001 年 1 月 1 日执行。按照《计算机信息系统安全保护等级划分准则》的规定，我国实行五级信息安全等级保护。

第一级：用户自主保护级。由用户来决定如何对资源进行保护，以及采用何种方式进行保护。

第二级：系统审计保护级。该级的安全保护机制支持用户具有更强的自主保护能力，特别是具有访问审记能力，即能创建、维护受保护对象的访问审计跟踪记录，记录与系统安全相关事件发生的日期、时间、用户和事件类型等信息。所有和安全相关的操作都能够被记录下来，以便当系统发生安全问题时，可以根据审计记录，分析追查事故责任人。

第三级：安全标记保护级。具有第二级系统审计保护级的所有功能，并对访问者及其访问对象实施强制访问控制。通过对访问者和访问对象指定不同安全标记，限制访问者的权限。

第四级：结构化保护级。将前三级的安全保护能力扩展到所有访问者和访问对象，支持形式化的安全保护策略。其本身构造也是结构化的，使之具有相当的抗渗透能力。该级的安全保护机制能够使信息系统实施一种系统化的安全保护。

第五级：访问验证保护级。具备第四级的所有功能，还具有仲裁访问者能否访问某些对象的能力。因此，该级的安全保护机制不能被攻击或篡改，具有极强的抗渗透能力。

思　考　题

(1) 简述网络安全的基本特征。

(2) 简述网络安全威胁的现状与发展趋势。

(3) ISO/OSI 安全体系有哪些内容？

(4) 简述 TCSEC 准则的网络安全分级情况。

参考文献

[1]　李仁发，等. 计算机网络安全. 北京：科学出版社，2004.
[2]　叶忠杰，等. 计算机网络安全技术. 北京：科学出版社，2003.
[3]　袁德明，乔月圆. 计算机网络安全. 北京：电子工业出版社，2007.
[4]　沈苏彬. 网络安全原理与应用. 北京：人民邮电出版社，2007.
[5]　徐建华，张英，万发仁. 影响网络安全的因素及防控措施. 农业网络信息，2008，(8)：80～82.
[6]　郑志彬. 信息网络安全威胁及技术发展趋势. 电信科学，2009，25(2)：28～34.

第 2 章　网络协议的安全

本章学习目标

TCP/IP 协议集确立了 Internet 的技术基础。本章分析 TCP/IP 协议对网络安全的影响以及常见的针对网络协议的攻击，并具体介绍网络层安全协议 IPSec、传输层安全协议 SSL、TLS 和应用层安全协议 SET、Telnet、HTTP 等。

通过对本章的学习，应掌握以下内容：

(1) TCP/IP 协议对网络安全的影响。

(2) 针对网络协议的攻击手段。

(3) IPSec 的安全技术。

(4) SSL 和 TLS 协议的基本原理。

(5) SET 协议的安全支付过程。

(6) Telnet 和 HTTP 的安全。

TCP/IP 协议是目前最常用的网络协议。在设计 TCP/IP 协议之初，设计者假设网络使用者是相互可信的，考虑得更多的是网络的互联互通，没有考虑其安全性，这导致 TCP/IP 协议存在致命的缺陷，例如，一些网络协议(Telnet、FTP 等)在对用户进行认证时，密码以明文的方式传输。

本章将分析 TCP/IP 协议对网络安全的影响，介绍常见的针对网络协议的攻击手段，最后对网络层、传输层以及应用层的网络安全协议进行介绍。

2.1　TCP/IP 协议与网络安全

2.1.1　TCP/IP 协议简介

TCP/IP 是一组 Internet 协议的总称，除常见的 TCP 协议和 IP 协议外，还包括其他协议，例如 FTP、UDP、ICMP 以及 Telnet 等。TCP/IP 协议的开发工作始于 20 世纪 70 年代，是用于 Internet 的第一套协议。TCP/IP 协议的开发研制人员将 Internet 分为 4 个层次，从下至上依次是数据链路层、网络层、传输层和应用层。

(1) 数据链路层：又称为链路层或网络接口层，对应于网络的基本硬件，这也是 Internet 的物理构成，即人们可以看得见的硬件设备，并定义了将数据组成正确帧的规程和在网络中传输帧的规程。例如 PC、服务器、网络设备等，必须对这些硬件设备的电气特性作一个规范，使这些设备都能够互相连接并兼容使用。

(2) 网络层：该层定义了网络中传输的信息包的格式，以及从一个用户通过一个或多个路由器到最终目标的信息包的转发机制。

(3) 传输层：为两个用户进程之间建立、管理和拆除可靠而又有效的端到端连接。

(4) 应用层：它定义了应用程序使用 Internet 的规程。

常见的 TCP/IP 协议如表 2-1 所示。

表 2-1　常见的 TCP/IP 协议列表

层　　名	TCP/IP 协议族
应用层	TFTP,HTTP,SNMP,FTP,SMTP,DNS,RIP,Telnet
传输层	TCP,UDP
网络层	IP,ICMP,OSPF,BGP,IGMP,ARP,RARP
数据链路层	SLIP,CSLIP,PPP,MTU,ARP,RARP,ISO 2110,IEEE 802,IEEE 802.2

以文件传输为例,TCP/IP 协议的工作原理为:

(1) 源主机应用层将相关数据流传输至传输层。

(2) 传输层将数据流进行分组,并加上 TCP 包头传输至网络层。

(3) 在网络层加上包括源、目的主机 IP 地址的 IP 报头,生成 IP 数据包,并将生成的 IP 数据包传输至链路层。

(4) 在链路层将 MAC 帧的数据部分装入 IP 数据包,然后将源、目的主机的 MAC 地址和帧头加上,并根据目的主机的 MAC 地址,将完整的 MAC 帧发往目的主机或者 IP 路由器。

(5) MAC 帧到达目的主机后,在链路层将 MAC 帧的帧头去掉,并将去掉 MAC 帧头的 IP 数据包传输至网络层。

(6) 网络层对 IP 报头进行检查,如果校验与计算结果不同,则将该 IP 数据包丢弃,如果结果一致就去掉 IP 报头,将 TCP 段传输至传输层。

(7) 传输层对顺序号进行检查,判断是否是正确的 TCP 分组,然后再对 TCP 报头数据进行检查。如果正确就向源主机发出确认信息,如果不正确或者是出现丢包,就向源主机发出重发要求。

(8) 在目的主机的传输层将 TCP 报头去掉后根据顺序对分组进行组装,然后将组装好的数据流传输至应用程序。

2.1.2　TCP/IP 协议的安全性

由于 TCP/ IP 协议在设计之初主要考虑让不同计算机之间、不同操作平台之间的通信成为可能,在当时的网络规模不大、应用范围不广、计算机技术尚不够发达的情况下,对安全问题考虑不多,因此 TCP/IP 协议在安全性方面做得不够完善。随着网络规模、计算机技术的日益发展,TCP/IP 协议的脆弱性日益突出,已开始阻碍 TCP/IP 协议的进一步广泛使用。

由于 TCP/IP 协议族自身存在一些安全缺陷,即使使用正确,TCP/IP 网络仍会受到攻击,例如序列号欺骗、路由攻击、源地址欺骗和授权欺骗等。

1. 数据链路层的安全性

在以太网中,数据是以帧为单位进行传输的。任何主机发送的帧都会到达与其处于同一网段的所有主机的网络接口,而每一个网络接口都有一个唯一的硬件地址,即网卡的 MAC 地址。信息以数据包的形式传输,其报头包含了目的主机的 MAC 地址,如果其携带的 MAC 地址是自己的或者是广播地址,那么就会将数据帧交给 IP 层,否则将其丢掉。目

前，网络上存在一些被称为嗅探器(Sniffer)的软件，例如 NeXRay、Sniffit 等，攻击者只要稍作设置或修改，使网卡工作在监听模式下，就可以达到非法窃取他人信息的目的。

2. 网络层的安全性

1) IP 协议的安全性

IP 协议是 TCP/IP 协议的核心，也是网络层中最重要的协议。IP 层接收低层发过来的数据包，并把该数据包发送到更高层的 TCP 层或 UDP 层；IP 层也把从 TCP 层或 UDP 层接收来的数据包传输到较低层。但是需要注意的是，IP 数据包是不可靠的，因为它无法确认数据是否按顺序发送和是否遭受破坏。IP 数据包中含有发送它的主机地址(源地址)和接收它的主机地址(目的地址)，高层 TCP 服务和 UDP 服务在接收数据包时，通常假设包中的源地址是有效的。目前，IP 地址已成为许多认证服务的基础，这些服务相信数据包是从一个有效的主机发送来的。IP 源路径是为了测试而存在的，它可以被用来欺骗系统进行平常被禁止的连接，因此许多依靠 IP 源地址做确认的服务将产生问题并且会被非法入侵。IP 源路径允许 IP 数据包自己选择一条通往系统目的主机的路径。设想攻击者试图与防火墙后面的一个不可到达的主机 A 连接。他只需要在送出的请求报文中设置 IP 源路径选项，使报文有一个目的地址指向防火墙，而最终地址是主机 A。当报文到达防火墙时被允许通过，因为它指向防火墙而不是主机 A，防火墙的 IP 层处理该报文的源路径被改变，并发送到内部网上，报文就这样到达了不可到达的主机 A。

2) ICMP 的安全

ICMP(Internet Control Message Protocol，Internet 控制报文协议)是 TCP/IP 协议族的一个子协议，用于在 IP 主机、路由器之间传输控制消息。利用 ICMP 漏洞可传输一些网络和主机的控制信息，例如目标主机是否可达、路由重定向等。常用的 ping 命令就使用了 ICMP 协议，它可通过发送一个 ICMP echo 请求消息和接收一个响应的 ICMP 回应来测试主机的连通性。通常也可以得到一些附加信息，例如收发数据包的往返时间等。几乎所有的 TCP/IP 机器都会对 ICMP echo 请求进行响应。

3) ARP 欺骗

ARP(Address Resolution Protocol，地址解析协议)欺骗即 ARP 重定向，就是向目标主机发送报文，其中含有攻击者伪造的 IP 地址和 MAC 地址，目标主机收到该报文后，会用报文中伪造的信息刷新 ARP 缓存。如果攻击者定时向目标主机发送该报文，而且时间间隔比 ARP 缓存的超时间隔小的话，目标主机就会一直维持着一张含有错误信息的 ARP 缓存表。

3. 传输层的安全性

TCP 会话劫持是传输层面临的一个重要安全威胁。TCP 会话劫持与 IP 欺骗不一样，IP 欺骗是针对 TCP 三次握手过程进行的攻击，而 TCP 会话劫持是跳过连接过程，对一个已经建立的连接进行攻击。TCP 通过三次握手建立连接以后，主要采用滑动窗口机制来验证对方发送的数据。如果对方发送的数据不在自己的接收窗口内，则丢弃此数据。这种发送序号不在对方接收窗口的状态称为非同步状态。当通信双方进入非同步状态后，攻击者可以伪造发送序号在有效接收窗口内的报文，也可以截获报文，篡改内容后再修改发送序号，而接收方会认为数据是有效数据。受这种攻击的主要原因是 TCP 协议并不对数据包进

行加密和认证，而是通过判断序列号是否正确来确认数据包。

4. 应用层的安全性

一些常用的应用层协议，例如 FTP、Telnet、HTTP 等由于自身不具备加密功能，导致其传输的敏感信息（例如密码等）很容易被其他人窃听到。在应用层常见的攻击手段是 DNS（Domain Name System，域名系统）欺骗。攻击者伪造计算机名称和网络的信息，当主机需要将一个域名转化为 IP 地址时，它会向某 DNS 服务器发送一个查询请求。同样，在将 IP 地址转化为域名时可发送一个反查询请求。如果服务器在进行 DNS 查询时人为地给出攻击者自己的应答信息，就产生了 DNS 欺骗。由于网络上的主机都信任 DNS 服务器，一个被破坏的 DNS 服务器就可以将客户引导到非法的服务器，从而就可以使某个地址产生欺骗。

2.2　针对网络协议的攻击

由上一节可知，TCP/IP 协议存在的漏洞可能会导致一些安全问题。针对网络协议的攻击有多种，比较典型的包括网络监听、拒绝服务攻击、TCP 会话劫持、网络扫描、重放攻击、数据修改以及伪装等。下面将具体介绍每一种针对网络协议的攻击特性。

2.2.1　网络监听

Ethernet（以太网）协议的工作方式是将要发送的数据包发往连接在一起的所有主机。在包头中包括有应该接收数据包主机的正确地址，因为只有与数据包中目标地址一致的那台主机才能接收到信息包，但是若主机工作在监听模式下，则不管数据包中的目标物理地址是什么，主机都将可以接收到。

通常，局域网内有十几台甚至上百台主机。它们通过一条电缆、一个集线器连接在一起。从协议的高层或者用户的角度来看，当同一网络中的两台主机通信时，源主机将写有目的主机地址的数据包直接发向目的主机，或者当网络中的一台主机同外界的主机进行通信时，源主机将写有目的主机 IP 地址的数据包发向网关。但这种数据包并不能在协议栈的高层直接发送出去，要发送的数据包必须从 TCP/IP 协议的 IP 层交给网络接口，即数据链路层。在网络接口由 IP 层来的带有 IP 地址的数据包又增加了一部分以太帧的帧头信息。在帧头中，有两个域分别为只有网络接口才能识别的源主机和目的主机的物理地址。物理地址是一个 48 位的地址，它与 IP 地址相对应，即一个 IP 地址和一个物理地址存在对应关系。对于作为网关的主机，由于它连接了多个网络，它也就同时具备有很多个 IP 地址，在每个网络中它都有一个。而发向网络外的帧中继携带的就是网关的物理地址。Ethernet 中填写了物理地址的帧从网络接口（网卡）中发送出去传输到物理线路上。如果局域网是由一条粗网或细网连接成的，那么数字信号在电缆上传输后就能够到达线路上的每一台主机。再当使用集线器的时候，发送出去的信号到达集线器，由集线器再发向连接在集线器上的每一条线路。这样在物理线路上传输的数字信号也就能到达连接在集线器上的每个主机了。当数字信号到达一台主机的网络接口时，正常状态下网络接口对读入数据帧进行检查，如果数据帧中携带的物理地址是自己的或者物理地址是广播地址，那么就会将数据帧交给 IP 层软件。

对于每个到达网络接口的数据帧都要进行这个过程。但是若主机工作在监听模式下,则所有的数据帧都将被交给上层协议软件处理。当连接在同一条电缆或集线器上的主机被逻辑地分为几个子网时,若有一台主机处于监听模式,它还将可以接收到发向与自己不在同一个子网(使用了不同的掩码、IP 地址和网关)的主机的数据包,在同一个物理信道上传输的所有信息都可以被接收到。

2.2.2 拒绝服务攻击

拒绝服务攻击(Denial of Service,DoS),即大家常说的 DoS 攻击。通常来说,凡是能导致合法用户不能进行正常的网络服务的行为都算是拒绝服务攻击。拒绝服务攻击的目的就是要阻止合法用户对网络资源的正常访问。从技术角度来看,拒绝服务就是用超出被攻击目标处理能力的海量数据包消耗可用系统、带宽资源,致使网络服务瘫痪的一种攻击手段。在早期,拒绝服务攻击主要针对处理能力比较弱的计算机或是带宽较小的网站,对拥有高带宽连接和高性能设备的网站影响不大。

在 1999 年底,伴随着分布式拒绝服务攻击(Distributed Denial of Service,DDoS)的出现,高端网站开始面临 DoS 的威胁。DDoS 实现是借助数百,甚至数千台被植入攻击守护进程的攻击主机同时发起的集体攻击行为。因此,DDoS 也被称为"洪水攻击"。常见的 DDoS 攻击手法有 UDP Flood、TCP SYN Flood、ICMP Flood 等。DDoS 攻击原理如图 2-1 所示。

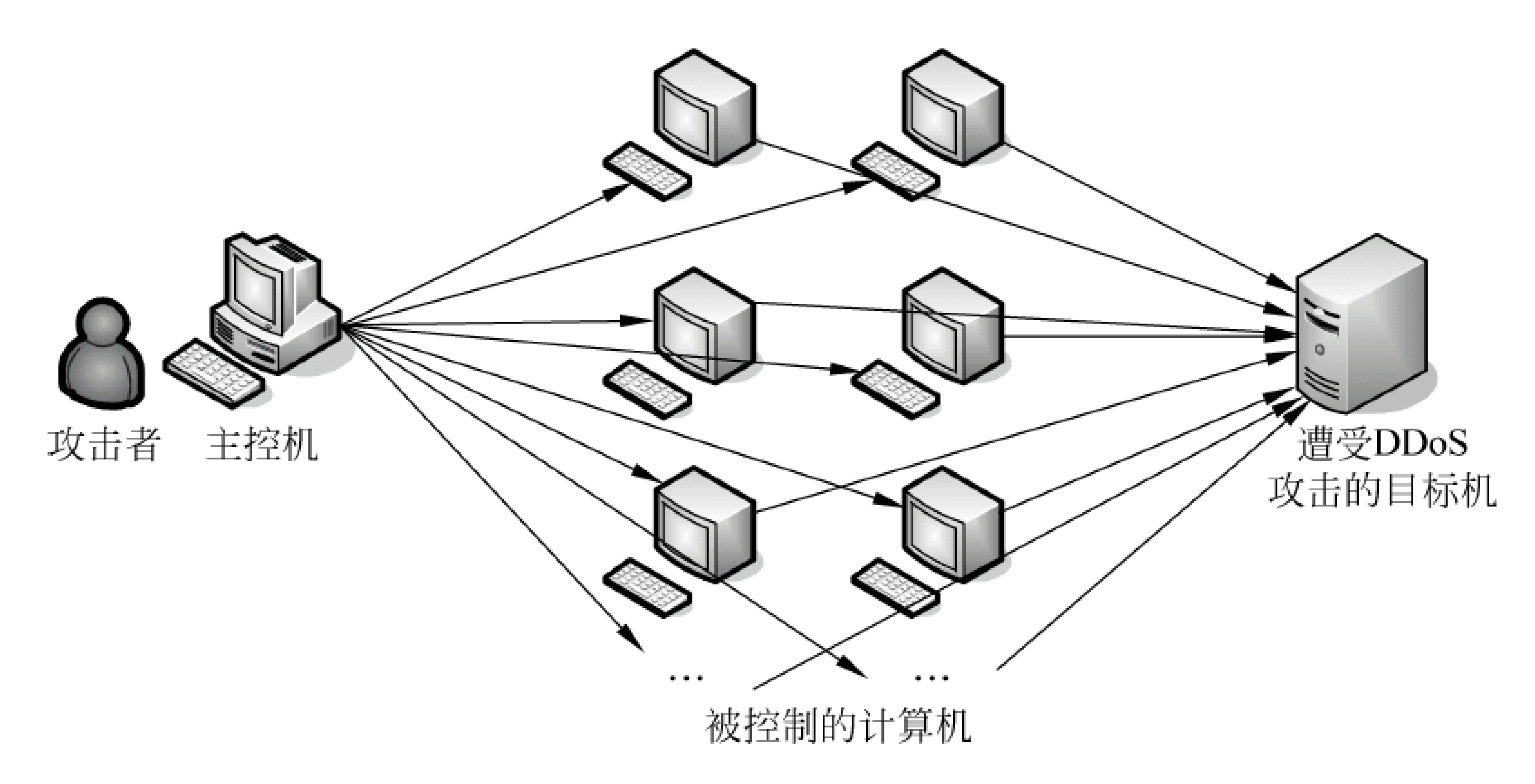


图 2-1 DDoS 攻击原理

1. UDP Flood

UDP Flood 是一种流量型 DoS 攻击,其原理是利用大量的 UDP 小包冲击 DNS 服务器或 RADIUS 认证服务器、流媒体视频服务器。在 UDP Flood 攻击中,攻击者可发送大量伪造源 IP 地址的小 UDP 包,但由于 UDP 协议是无连接性的,所以只要开通了 UDP 端口提供相关服务的话,就可针对相关的服务进行攻击。正常应用中,UDP 包双向流量基本相等,而且大小和内容都是随机的,变化很大。在出现 UDP Flood 的情况下,针对同一目标 IP 的 UDP 包在一侧大量出现,并且内容和大小都比较固定。

2. TCP SYN Flood

在 TCP 连接的三次握手中，假设一个用户向服务器发送了 SYN 报文后突然死机或掉线，那么服务器在发出 SYN＋ACK 应答报文后无法再收到客户端的 ACK 报文，即第三次握手无法完成。此时服务器端一般会重试(即再次发送 SYN＋ACK 给客户端)并等待一段时间，超时后丢弃这个未完成的连接，称这段时间的长度为 SYN Timeout，时间的数量级是分钟(通常为 30 秒～2 分钟)。在实际应用中，一个用户出现异常导致服务器的一个线程等待 1 分钟并不是大问题，但若存在一个恶意的攻击者大量模拟这种情况，服务器端将为了维护一个非常大的半连接列表而消耗非常多的资源，大量的 CPU 时间和内存被占用。如果服务器的 TCP/IP 栈不够强大，往往会导致堆栈溢出崩溃；即使服务器端的系统足够强大，服务器端也将忙于处理攻击者伪造的 TCP 连接请求而无暇理睬客户的正常请求。此时从正常客户的角度看来，服务器失去响应，这种情况称为 TCP SYN Flood 攻击。

3. ICMP Flood

ICMP 的控制消息是指网络通不通、主机是否可达、路由是否可用等网络本身的消息。这些控制消息虽然并不传输用户数据，但是对于用户数据的传输起着重要作用。

由于早期阶段的路由器对数据包的最大尺寸都有限制，许多操作系统对 TCP/IP 的实现在 ICMP 包上都规定是 64KB。ICMP Flood 是一种 DDoS 攻击，通过对其目标大量发送超过 65 535 字节(64KB)的数据包，就会出现内存分配错误，从而可以导致目标主机瘫痪。

2.2.3　TCP 会话劫持

TCP 是一个可靠的、连接定向的发送服务，数据分段传输。连接定向意味着在主机交换数据之前必须建立会话。TCP 使用字节流通信，这意味着数据被当作没有边界的字节序列。TCP 会话的可靠性通过给传输的段分配序号来实现。如果一个 TCP 段被分为若干小片，接收主机知道是否所有的片都收到了，并通过答复的方法检验数据是否被其他主机接收到。对于每一个发送的段，接收主机必须返回带有特定时间段内接收到的字节数的答复(ACK)。如果没接收到 ACK，数据将重新发送；如果段在接收到的时候已经损坏，接收主机将废弃它。因为没有发回 ACK，所以发送者将重新发送这段。

一次 TCP 会话会通过三次握手来实行初始化。三次握手过程包括：

(1) 源主机通过发送带有置为 on 的 SYN 标志的段发送会话请求。

(2) 接收主机通过发回具有以下特征的数据段表示同意接收：SYN 标志置为 on，即将发送数据段的起始字节的序列号、应答和带有它等待接收的下一个数据段的字节序列号。

(3) 请求会话的主机再回送一个数据段，并带有确认序列号和确认号。

TCP 使用类似的握手过程结束连接，以保证两台主机都结束传输并且所有的数据收到。

根据 TCP/IP 的规定，使用 TCP 协议进行通信需要提供两段序列号，TCP 协议使用这两段序列号确保连接同步以及安全通信，系统的 TCP/IP 协议栈依据时间或线性地产生这些值。在通信过程中，双方的序列号是相互依赖的。如果攻击者在这个时候进行会话劫持，结果肯定是失败，因为会话双方“不认识”攻击者，攻击者不能提供合法的序列号。所以，会话劫持(Session Hijacking)的关键是预测正确的序列号，攻击者可以采取嗅探技术获得这

些信息。

1. TCP 协议的序列号

在每一个数据包中都有两段序列号，它们分别为：

(1) SEQ：当前数据包中的第一个字节的序号。

(2) ACK：期望收到对方数据包中的第一个字节的序号。

假设双方现在需要进行一次连接，服务器端和客户端分别具有如下参数：

(1) 服务器端：

S_SEQ：将要发送的下一个字节的序号。

S_ACK：将要接收的下一个字节的序号。

S_WIND：接收窗口。

(2) 客户端：

C_SEQ：将要发送的下一个字节的序号。

C_ACK：将要接收的下一个字节的序号。

C_WIND：接收窗口。

这些参数之间必须符合如下的逻辑关系，否则该数据包会被丢弃，并且返回一个 ACK 包(包含期望的序列号)。

(1) C_ACK≤C_SEQ≤C_ACK + C_WIND。

(2) S_ACK≤S_SEQ≤S_ACK + S_WIND。

如果不符合上边的逻辑关系，就会引申出一个“致命弱点”。

2. 致命弱点

这个致命的弱点就是 ACK 风暴。当会话双方接收到一个不期望的数据包后，就会用自己期望的序列号返回 ACK 包；而在另一端，这个数据包也不是所期望的，就会再次以自己期望的序列号返回 ACK 包。于是就这样来回往返，形成了一个恶性循环，最终导致了 ACK 风暴。比较好的解决办法是先进行 ARP 欺骗，使双方的数据包“正常”地发送到攻击者这里，然后设置包转发，最后就可以进行会话劫持了，而且不必担心会有 ACK 风暴出现。当然，并不是所有系统都会出现 ACK 风暴，例如 Linux 系统的 TCP/IP 协议栈就与 RFC 中的描述略有不同。

3. TCP 会话劫持过程

假设现在主机 A 和主机 B 进行一次 TCP 会话，C 为攻击者，劫持过程如下(其中，X、Y 代表不同的随机序列号；FLAG 表示标识；AP 表示 TCP 的标识；ACK PUSH 表示把确认信息(ACK)强制转交高层；Window：ZZZZ 表示发送的包数据，包的大小不需要加单位。)：

(1) A 向 B 发送一个数据包：

```
SEQ (hex): X   ACK (hex): Y
FLAGS: - AP---  Window: ZZZZ,包大小为 60
```

(2) B 回应 A 一个数据包：

```
SEQ (hex): Y   ACK (hex): X + 60
FLAGS: - AP---  Window: ZZZZ,包大小为 50
```

(3) A 向 B 回应一个数据包:

```
SEQ (hex): X + 60   ACK (hex): Y + 50
FLAGS: - AP ---  Window: ZZZZ,包大小为 40
```

(4) B 向 A 回应一个数据包:

```
SEQ (hex): Y + 50   ACK (hex): X + 100
FLAGS: - AP ---  Window: ZZZZ,包大小为 30
```

(5) 攻击者 C 冒充主机 A 给主机 B 发送一个数据包:

```
SEQ (hex): X + 100   ACK (hex): Y + 80
FLAGS: - AP ---  Window: ZZZZ,包大小为 20
```

(6) B 向 A 回应一个数据包:

```
SEQ (hex): Y + 80   ACK (hex): X + 120
FLAGS: - AP ---  Window: ZZZZ,包大小为 10
```

现在,主机 B 执行了攻击者 C 冒充主机 A 发送过来的命令,并且返回给主机 A 一个数据包;但是主机 A 并不能识别主机 B 发送过来的数据包,所以主机 A 会以期望的序列号返回给主机 B 一个数据包,随即形成 ACK 风暴。如果成功地解决了 ACK 风暴,就可以成功进行 TCP 会话劫持了。

2.2.4　网络扫描

安全扫描也称为脆弱性评估,它是检测远程或本地系统安全脆弱性的一种安全技术。其基本原理是采用模拟黑客攻击的方式对目标可能存在的已知安全漏洞进行逐项检测,以便对工作站、服务器、交换机、数据库等各种对象进行安全漏洞检测。借助于扫描技术,人们可以发现网络和主机存在的对外开放的端口、提供的服务、某些系统信息、错误的配置、已知的安全漏洞等。因此安全扫描技术是一种极为有效的主动防御技术,结合入侵检测系统和防火墙等其他安全技术,可为网络提供全方位的保护。

目前主要的安全扫描技术包括端口扫描技术、操作系统检测技术以及漏洞扫描技术。

1. 端口扫描技术

TCP 协议和 UDP 协议是 TCP/IP 协议传输层中两个用于控制数据传输的协议。TCP 和 UDP 用端口号来唯一地标识一种网络应用。TCP 和 UDP 端口号用 16 位二进制数表示,理论上每一个协议可以拥有 65 535 个端口。因此,端口扫描无论是对网络管理员还是对网络攻击者来说都是非常重要的。

TCP/IP 协议上的端口有 TCP 端口和 UDP 端口两类。由于 TCP 协议是面向连接的协议,针对 TCP 扫描方法比较多,从最初的一般探测发展到后来的躲避 IDS 和防火墙的高级扫描技术。针对 TCP 端口的扫描最早出现的是全连接扫描,随着安全技术的发展,出现了以躲避防火墙为目的的 TCP SYN 扫描以及其他一些秘密扫描技术,例如 TCP FIN 扫描、TCP ACK 扫描、NULL 扫描、XMAS 扫描、SYN/ACK 扫描和 Dumb 扫描等。UDP 端

口的扫描方法相对比较少,只有 UDP ICMP 端口不可达扫描和利用 socket 函数 xecvfrom 和 write 来判断的扫描。目前,端口扫描技术已经发展得非常丰富和完善。端口扫描主要可分为开放扫描、半开放扫描、秘密扫描等。

1) TCP connect 扫描

这是最基本的 TCP 扫描方法。使用操作系统提供的 connect( )系统调用来与目标主机的 TCP 端口进行连接。如果 connect( )连接成功,说明目标端口处于监听状态。若连接失败,则说明该端口没有开放。TCP connect( )方法的最大优点是无须任何特殊权限,系统中任何用户都可以调用 connect( )函数。还有一个优点就是速度快。但 TCP connect( )方法的缺点是它很容易被发觉,并且容易被防火墙过滤掉;同时目标主机的日志文件会记录一系列的有关该服务的连接建立并马上断开的错误信息。

2) TCP SYN 扫描

该技术通常称为半开放(或半连接)的扫描,这是因为扫描程序不需要建立一个完全的 TCP 连接。扫描程序发送的是一个 SYN 数据包,就像准备建立一个实际的连接并等待回应一样。若返回 SYN ACK 数据包则表示目标端口处于监听状态,而若返回 RST 数据包则表示该端口没有开放。如果收到 SYN ACK 数据包,则扫描程序再发送一个 RST 数据包来终止该连接过程。TCP SYN 扫描技术的优点是一般不会在目标主机上留下记录。而该方法的缺点是必须要有管理员权限,自己构造的 SYN 数据包才能使用这种扫描方式。

3) TCP FIN 扫描

对于 TCP SYN 扫描,有些防火墙和包过滤系统能监视并限制可以接收 SYN 的端口,并能检查到这些扫描。对于 FIN 数据包则没有任何麻烦,可以顺利地通过。这种扫描方法的思想是不开放的端口会用 RST 来应答 FIN 数据包,而开放的端口会忽略对 FIN 数据包的应答。该方法和系统的实现相关,有的系统不管端口是否开放,都回复 RST 包。这种情况下该扫描方法就不适用了。

4) TCP 反向 ident 扫描

ident 协议(RFC 1413)可以获取任何使用 TCP 连接进程的运行用户名,即使该进程没有发起连接。例如用户连接到 HTTP 端口,然后用 identd 来获取 HTTP 服务是否以管理员用户的身份在运行。该方法需要和目标端口建立一个完整的 TCP 连接后方能使用。

5) FTP 返回攻击扫描

FTP 协议(RFC 959)支持代理 FTP 连接,即可以从本地计算机连接到目标主机 FTP 协议解释器,以建立控制连接。向目标主机的协议解释器发送建立数据传输进程的请求就可以请求 FTP 服务器向 Internet 上任何地方发送文件。该扫描方法就是使用 port 命令来声明本地的客户数据传输进程正在被动地在扫描目标主机的某个端口监听。然后再用 list 命令请求列出当前目录,服务器将结果通过数据传输进程发送到指定扫描目标主机的端口。如果目标主机正在该端口监听,传输就会成功(产生一个 150 或 226 的应答)。否则,会出现连接被拒绝错误(426 应答)。这样就探测到目标主机端口是否开放的信息。这种方法的优点是它不容易被追踪,并可能穿过防火墙;缺点是速度很慢,而且有些 FTP 服务器能发现这种扫描而关闭代理功能。因此该方法只能适用于部分 FTP 服务器。

2. 操作系统探测技术

1）应用层探测技术

通过向目标主机发送应用服务连接或访问目标主机开放的有关记录就可能探测出目标主机的操作系统(包括相应的版本号)。

2）TCP/IP 堆栈特征探测技术

目前流行的 TCP/IP 堆栈特征探测技术有：

(1) FIN 探测：通过发送一个 FIN 数据包到一个打开的端口，并等待回应。RFC 793 定义的标准行为是“不”响应，但 Windows、BSD、Cisco 等操作系统会回应一个 RESET 包。大多数的探测器都使用了这项技术。

(2) BOGUS 标记位探测：通过发送一个 SYN 包，它含有没有定义 TCP 标记的 TCP 头。那么在 Linux 系统的回应中仍旧会包含这个没有定义的标记，而在一些别的系统则会在收到该包之后关闭连接。利用这个特性可以区分一些操作系统。

(3) TCP ISN 取样：这是利用寻找初始化序列规定长度与特定的操作系统相匹配的方法。利用它可以对许多系统分类，例如较早的 UNIX 系统是 64KB 长度。一些新的 UNIX 系统则是随机增长的长度，而 Windows 平台则使用基于时间方式产生的 ISN 会随着时间的变化而有着相对固定的增长。

3. 漏洞扫描技术

漏洞扫描就是通过采用一定的技术主动地去发现系统中的安全漏洞。漏洞扫描可以分为对未知漏洞的扫描和对已知漏洞的扫描。未知漏洞扫描的目的在于发现软件系统中可能存在但尚未发现的漏洞。已知漏洞的扫描主要是通过采用模拟黑客攻击的方式对目标可能存在的已知安全漏洞进行逐项扫描。漏洞扫描技术是建立在端口扫描技术和远程操作系统识别技术的基础之上的。漏洞扫描主要通过以下两种方法来检查目标主机是否存在漏洞：

1）特征匹配方法

基于网络系统漏洞库的漏洞扫描的关键部分就是它所使用的漏洞特征库。通过采用基于规则的模式特征匹配技术，即根据安全专家对网络系统安全漏洞、黑客攻击案例的分析和系统管理员对网络系统安全配置的实际经验，可以形成一套标准的网络系统漏洞库，然后再在此基础之上构成相应的匹配规则，由扫描程序自动进行漏洞扫描。若没有被匹配的规则，系统的网络连接是禁止的。

2）插件技术

插件是由脚本语言编写的子程序，扫描程序可以通过调用它来执行漏洞扫描，检测出系统中存在的一个或多个漏洞。添加新的插件就可以使漏洞扫描软件增加新的功能，扫描出更多的漏洞。

2.2.5　重放攻击

重放攻击(Replay Attacks)又称重播攻击、回放攻击或新鲜性攻击(Freshness Attacks)。它的基本原理是指攻击者将目的主机已接收过的包重新发送给接收方来达到欺骗系统的目的。重放攻击主要用于身份认证过程，用来破坏认证的正确性。重放攻击可以由发起者，也可以由拦截并重发该数据的敌方进行。攻击者利用网络监听或者其他方式盗

取认证凭据，之后再把它重新发给认证服务器。根据重放攻击的原理，虽然采用加密的方法可以有效防止会话劫持，但是却无法防止重放攻击。重放攻击是计算机世界黑客常用的攻击方式之一。

2.2.6 数据修改

攻击者读取数据之后，可以在发送者或接收者未察觉的情况下修改数据包中的数据。对于通信过程中传输的数据，无论是否加密，使用者都不希望任何信息被修改。

2.2.7 伪装

严格来说，伪装攻击也属于数据修改的范畴，不同的是伪装攻击修改的是 IP 地址。由于 IP 地址在网络中为计算机身份的有效标识，因此攻击者采用特殊程序构造 IP 数据包、使数据包看起来是来自内部网中的有效地址。

伪装具有多种用途。例如攻击者可以通过使用假的 IP 地址隐藏自己的身份，以逃避检测人员的追踪，同时攻击者可以通过它进行 DoS 攻击。此外，攻击者通过 IP 伪装欺骗用户验证机制，冒充合法并进行非法活动等。

2.3 网络层的安全

网络攻击的根源是没有对所传输的数据进行加密，且对传输两端缺少可靠的身份验证。这些功能可以在 TCP/IP 模型的不同层实现。在网络层上的解决方案是 IPSec 协议。

IPSec 协议不是一个单独的协议，它给出了应用于 IP 层上网络数据安全的一整套体系结构，包括网络认证协议（Authentication Header，AH）、封装安全载荷协议（Encapsulating Security Payload，ESP）、密钥管理协议（Internet Key Exchange，IKE）和用于网络认证及加密的一些算法等。IPSec 规定了如何在对等层之间选择安全协议、确定安全算法和密钥交换，向上提供了访问控制、数据源认证、数据加密等网络安全服务。

2.3.1 IPSec 的安全特性

IPSec 的安全特性主要包括：

（1）不可否认性。可以证实消息发送方是唯一可能的发送者，且发送者不能否认发送过消息。这里采用了公钥密码技术的一个特征，当使用公钥密码技术时，发送方用私钥产生一个数字签名随消息一起发送，接收方用发送者的公钥来验证数字签名。由于在理论上只有发送者才唯一拥有私钥，也只有发送者才可能产生该数字签名，所以只要数字签名通过验证，发送者就不能否认曾发送过该消息。但不可否认性不是基于认证的共享密钥技术的特征，因为在基于认证的共享密钥技术中，发送方和接收方掌握相同的密钥。

（2）抗重播性。抗重播性即确保每个 IP 包的唯一性，保证信息万一被截取复制后，不能再被重新利用、重新传输回目的地址。该特性可以防止攻击者截取破译信息后，再用相同的信息包冒取非法访问权。

（3）数据完整性。防止传输过程中数据被篡改，确保发出数据和接收数据的一致性。

IPSec 利用 Hash 函数为每个数据包产生一个加密检查和,接收方在打开包前先计算检查和,若包遭篡改导致检查和不相符,数据包即被丢弃。

(4) 数据保密性。在传输前先对数据进行加密,可以保证在传输过程中即使数据包遭截取,信息也无法被读取。该特性在 IPSec 中为可选项,与 IPSec 策略的具体设置相关。

(5) 身份认证。数据源发送信任状,由接收方验证信任状的合法性,只有通过认证的系统才可以建立通信连接。

2.3.2 IPSec 的体系结构

IPSec 的体系结构如图 2-2 所示。它显示了 IPSec 的总体结构、组成部件以及各部件之间的相关关系。IPSec 组建包含安全协议验证头 AH、封装安全载荷 ESP、安全关联(Security Association,SA)、密钥管理 IKE 以及加密算法和验证算法等。其中,加密算法和验证算法是 IPSec 实现安全数据传输的核心。

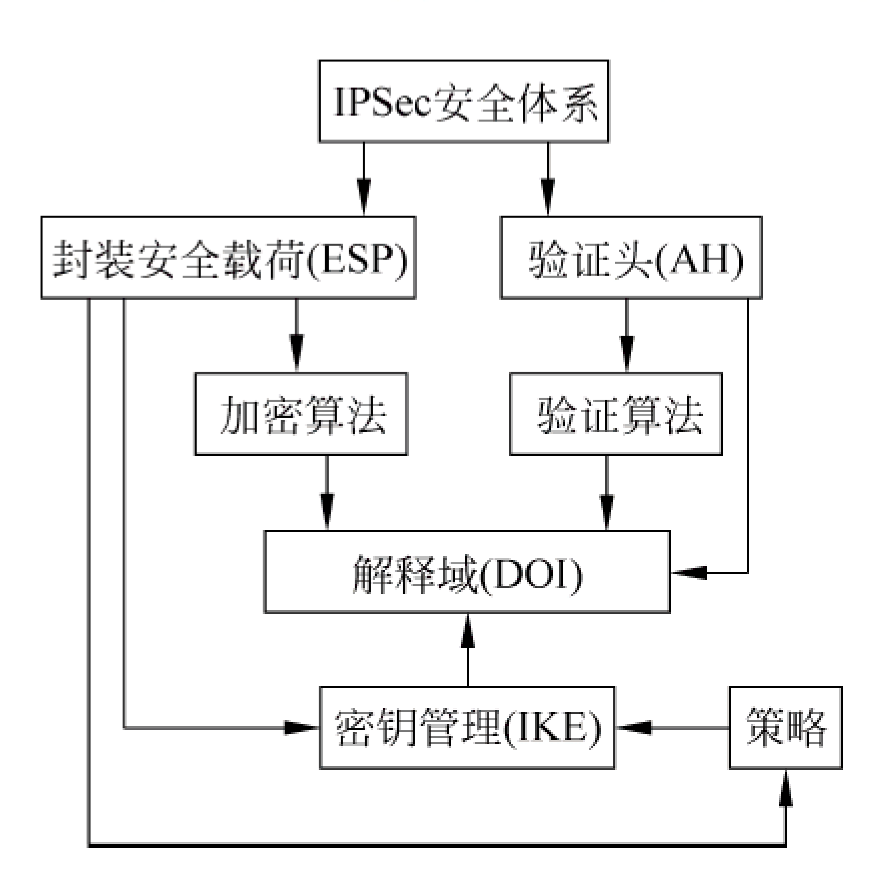


图 2-2 IPSec 的体系结构

2.3.3 AH 协议

IP 协议中,用来提供 IP 数据包完整性的认证机制是非常简单的。IP 头通过头部的校验和域来保证 IP 数据包的完整性。而校验和只是对 IP 头的每 16 位计算累加和的反码。这样并没有提供多少安全性,因为 IP 头很容易被修改,可以对修改过的 IP 头重新计算校验和并用它代替以前的校验和。这样接收端的主机就无法知道数据包已经被修改。

设计认证头(AH)协议的目的是用于增加 IP 数据包的安全性。AH 协议提供无连接的完整性(Connectionless Integrity)、数据源认证(Data Origin Authentication)和抗重播(Anti-Replay)攻击服务(在 IPSec 中,这三项功能混合在一起称为认证)。数据完整性是通过消息认证码生成的校验值来保证的;数据源认证是通过在数据包中含有一个将要被认证的共享密钥来保证的;而抗重放攻击是通过在 AH 中使用一个序列号来实现的。然而,AH 不对 IP 包提供任何保密性服务,也就是说它不加密所保护的数据包。AH 的作用是为 IP 数据流提供高强度的密码认证,以确保被修改过的数据包可以被检查出来。AH 使用消息认证码(Message Authentication Code,MAC)对 IP 进行认证。MAC 不同于 Hash 函数,因为它需要密钥来产生消息摘要,而 Hash 函数不需要密钥。常用的 MAC 是 HMAC,它与任何迭代密码 Hash 函数(例如 MD5、SHA-1、Tiger 等)结合使用,而不用对 Hash 函数进行修改。

AH 定义保护方法、头的位置、身份验证的覆盖范围以及输出和输入处理规则,但没有对所用的身份验证算法进行定义。AH 也没有硬性规定抗重播保护,使用抗重播服务由接收端自行处理,因而发送端无法得知接收端是否会检查其序列号。其结果是,发送端则一直认定接收端正在使用抗重播服务。

1. AH 的格式

AH 被分配到的标识数为 51,这表示 AH 保护的 IP 包的协议字段值为 51。AH 头紧

跟在IPv4/IPv6报头之后,格式如图2-3所示。

下一个荷载头	载荷长度	保留
安全参数索引(SPI)		
序列号		
认证数据(可变长)		

图2-3　AH的格式

(1) 下一个荷载头(Next Header):8bit字段,表示AH下一个载荷的协议类型。

(2) 载荷长度(Payload Length):8bit字段,AH的载荷长度,以32bit字为单位的认证头的长度减去2。

(3) 保留(Reservation):8bit字段,保留供将来使用。

(4) 安全参数索引(SPI):它是一个32bit长的整数字段。它与源地址或目的地址以及AH来共同唯一标识一个数据包所属的数据流的安全关联(SA)。SPI的值1～255被IANA留作将来使用,0被保留用于本地和具体实现。所以目前有效的SPI值从256～$2^{32}-1$。

(5) 序列号(Sequence Number):这里包含了一个作为单调增加计数器的32bit无符号整数字段,用于防止对数据包的重播。如果接收端启动了反重播攻击功能,它将使用滑动接收窗口检测重放数据包。

(6) 认证数据(Authentication Data):这是一个可变长的字段(必须是32bit的整数倍)。它包含数据包的认证数据,该认证数据被称为这个数据包的完整性校验值(ICV)或MAC。用于计算ICV的可用的算法因IPSec实现的不同而不同;然而,为了保证互操作性,AH强制所有的IPSec必须包含两个MAC:HMAC-MD5－96和HMAC-SHA-I－96。

2. AH协议的两种模式

AH可以用于两种模式:传输模式和隧道模式,如图2-4所示。在这两种模式下,AH都要对外部IP头的固有部分进行身份验证。

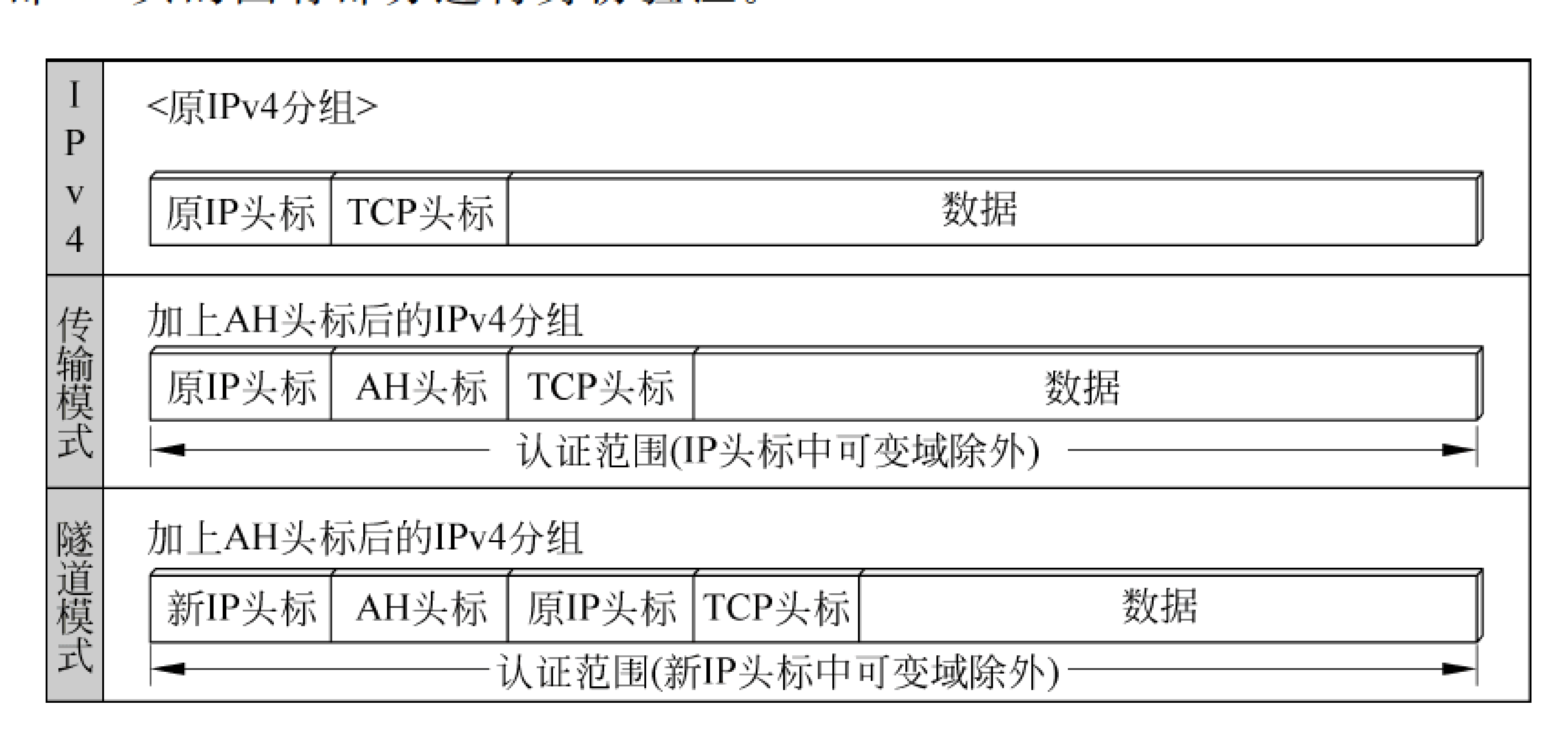


图2-4　IPv4中AH的传输模式与隧道模式

1) 传输模式

AH用于传输模式时,保护的是端到端的通信,通信的终点必须是IPSec终点。AH头被插在数据包中,紧跟在IP头之后和需要保护的上层协议之前,对这个数据包进行安全

保护。

2）隧道模式

AH 用于隧道模式时，需要将自己保护的数据包封装起来，并且 AH 头之前另外添了一个 IP 头，对整个 IP 数据包提供认证保护。“里面的”原 IP 数据包中包含了通信的原始寻址，而“外面的”新 IP 数据包则包含了 IPSec 端点的地址。隧道模式可用来替换端到端安全服务的传输模式，但是由于这一协议中没有提供机密性，因此相应地就没有通信分析这一保护措施，所以它没什么用处。AH 只用于保证收到的数据包在传输过程中不会被修改，保证由要求发送它的当事人将它发送出去，以及保证它是一个新的非重放的数据包。

3. AH 的处理

1）输出处理

对于输出的数据包，AH 协议处理的目标是向数据包合适的位置增加 AH 报头。具体的输出处理步骤如下：

（1）外出数据包与一个 SPDB 条目匹配时，查看 SADB 是否有合适的 SA。如果有，就将 AH 应用到与这个与之相符的数据包，该数据包在 SPDB 条目指定的那个模式中。如果没有，要么手工，要么通过 IKE 动态地建立一个，并且把序列号计数器初始化为 0。在利用这个 SA 构建一个 AH 头之前，计数器就开始递增，这样保证了每个 AH 报头中的序列号都是一个独一无二的单向递增的非零数。

（2）将 AH 的其余字段填满恰当的值。SPI 字段分配的值是取自 SA 的 SPI，下一个荷载头字段分配的是跟在 AH 之后的数据类型值，而载荷长度分配的则是 32 位字减 2；认证数据字段则设成 0。需要注意的是：AH 协议将安全保护扩展到外部 IP 包头的原有的或预计的字段，因此将完整性检查值（ICV）之前的不定字段清零是必要的。

（3）根据验证算法的要求，或出于排列方面的原因，需要进行适当的填充。对有些 MAC 算法来说，例如 DES-CBC MAC，要求应用 MAC 的数据必须是算法的块尺寸大小的倍数。在这种情况下就必须进行填充以便正确地使用 MAC。填充的数据包必须为零，并且填充数据的长度不包括在载荷长度中。对 IPv4 来说 AH 报头必须是 32bit 的倍数，IPv6 则是 64bit 的倍数。如果 MAC 算法的输出不符合这项要求就必须添加 AH 报头。对填充项的值没有什么别的要求，但必须把它包括在 ICV 的计算中，而载荷长度中必须反映出填充项大小。

（4）计算 ICV。从输出 SA 中取出验证密钥，连同整个 IP 包（包括 AH 报头）传到特定的算法（即 SA 中的身份验证程序）计算 ICV。由于不定字段已清零，它们不会被包括在 ICV 的计算中。将计算得到的 ICV 值复制到 AH 的认证数据字段中，IP 包头中的不定字段就可根据 IP 处理的不同得以填充。

（5）输出已处理的报文。AH 处理结束后就形成了 AH 保护下的 IP 数据包，根据数据包的大小，在传输到网络上前可将它分段处理，或在两个 IPSec 同级之间的传输过程中，由路由器分段。

2）输入处理

对于输入的数据包，AH 协议处理的目标就是从数据包中将 AH 报头剥离下来，还原出封装在 IPSec 内的高层数据包，其具体处理过程如下：

（1）重组分段。如果一个受 AH 安全保护的包在接收时被证实是分段数据，那么在

AH 输入处理之前需要对这些分段数据进行重新组合。因为如果分段的数据包没有重组为原来的完整数据,ICV 检查就会失败。只有完整的 AH 保护的 IP 包可传输到 AH 输入处理。

(2) 查询 SADB,找出保护这个包的 SA。用基于 IP 包头的 SPI、目的 IP 地址和安全协议(AH)组成的三元组来对 SA 进行查询。如果没有找到合适的 SA,这个包将会被丢弃。

(3) 进行序列号检查。如果检查失败,这个包就会被丢弃。在这个过程中的抗重放检查会决定这个包是新收到的还是以前收到的。

(4) 检查 ICV。首先把 AH 报头中的认证数据字段中的 ICV 值取出来,然后将这个字段清零,同时将 IP 中所有不定字段也清零。根据验证算法的要求以及载荷长度的要求可能还要进行零数据的填充,使验证数据的长度符合算法的要求。随后对整个数据包应用验证算法,并将获得的摘要同保存下来的 ICV 值进行比较。若相符,IP 包就通过了身份验证;否则将该数据包丢弃。

(5) 接收窗口的序列号可以递增,结束 AH 处理过程。验证通过的整个数据包传递给下一步的 IP 来处理。

2.3.4 ESP 协议

由于认证只确认了信息包的来源和完整性,而不能保护内容的机密性,为此需要引入机密性服务机制 ESP。封装安全载荷协议 ESP 为 IP 报文以无连接的方式(以包为单位)提供完整性校验、认证和加密服务,同时还可能提供防重放攻击保护和流量控制等服务。在建立 SA 时可选择所期望得到的安全服务,建议遵守以下约定:

(1) 完整性校验和身份认证建议同时使用。

(2) 使用防重放攻击时建议同时使用完整性校验和身份认证。

(3) 防重放攻击保护的使用建议由接收端选择。

(4) 加密独立于其他的安全服务,但建议在使用加密时同时使用完整性校验和身份认证。

1. ESP 协议包格式

ESP 协议隧道模式的包格式如图 2-5 所示。

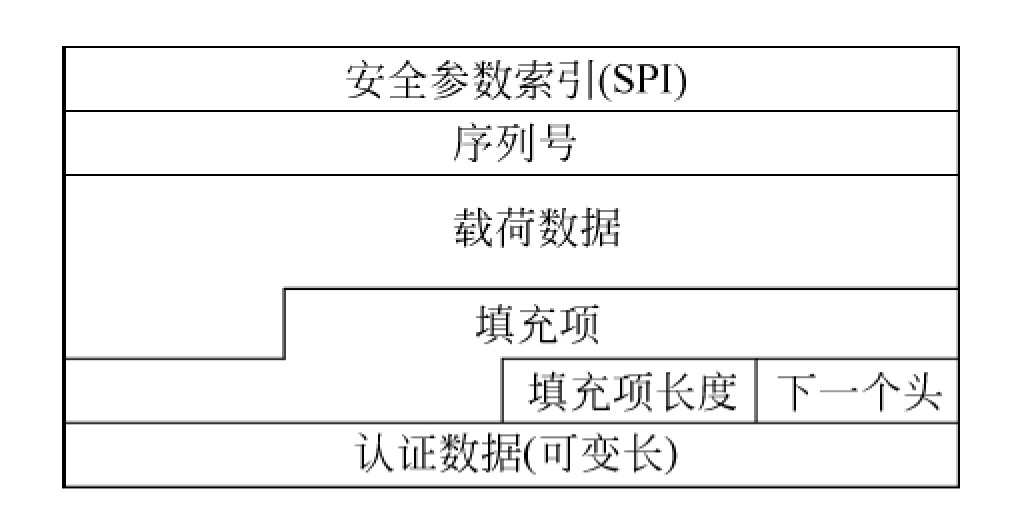


图 2-5 ESP 头(尾)的格式

(1) 安全参数索引(SPI):32bit 字段,与目的 IP 地址和 ESP 结合在一起,用来标识处理数据包所属的安全关联 SA。SPI 一般在 IKE 交换过程中由目标主机选定,当其值为 0 时,表示预留给本地使用。

(2) 序列号(Sequence Number):32bit 字段,是一个单项递增的计数器,用于防止重放攻击。无论接收者是否选择使用特定 SA 的抗重放服务,都必须使用序列号,并由接收者选择是否需要处理序列号字段。当建立一个 SA 时,发送者和接收者的计数器初始化为 0,并在进行 IPSec 输出处理前,使这个值递增。新的 SA 必须在序列号归 0 之前创建。

(3) 载荷数据(Payload Data)：可变长字段，是 ESP 保护的实际数据包，数据类型由下一载荷头字段来表示。在这个域中，包含下一个头字段，也可包含一个加密算法可能需要使用到的初始化向量(Initialization Vector，IV)，虽然载荷数据是加密的，但 IV 是没有加密的。

(4) 填充项(Padding)：字段范围为 0～255 字节，填充项的使用是为了保证 ESP 的边界适合于加密算法的需要。因为有些加密算法要求输入数据是以一定数量的字节为单位的块的整数倍数，即使 SA 没有机密性要求，仍然需要通过加入填充数据把 ESP 报头的填充项长度和下一个头这两个字段靠右排列。

(5) 填充项长度(Pad Length)：指出上面的填充项填充了多少字节的数据，因此，接收端可以恢复出载荷数据的真实长度。

(6) 下一个头(Next Header)：8bit 字段，表明包含在载荷数据字段的类型，字段的大小从 IP 协议数据中选择。在隧道模式下使用 ESP，这个值是 4，表示 IP-in-IP。如果在传输模式下使用 ESP，这个值就表示它背后的上一级协议的类型。

(7) 认证数据(Authentication Data)：该字段的长度由选择的认证功能指定。它包含数据完整性检验结果(Integrity Check Value，ICV)。验证数据计算的是 ESP 包中除验证数据域以外的所有项。如果对 ESP 数据包进行处理的 SA 中没有指定身份验证器，就没有这一项。

2. ESP 传输模式

同 AH 一样，ESP 也可提供两种模式的服务：传输模式和隧道模式，如图 2-6 所示，二者的差别决定了 ESP 保护的真正对象。

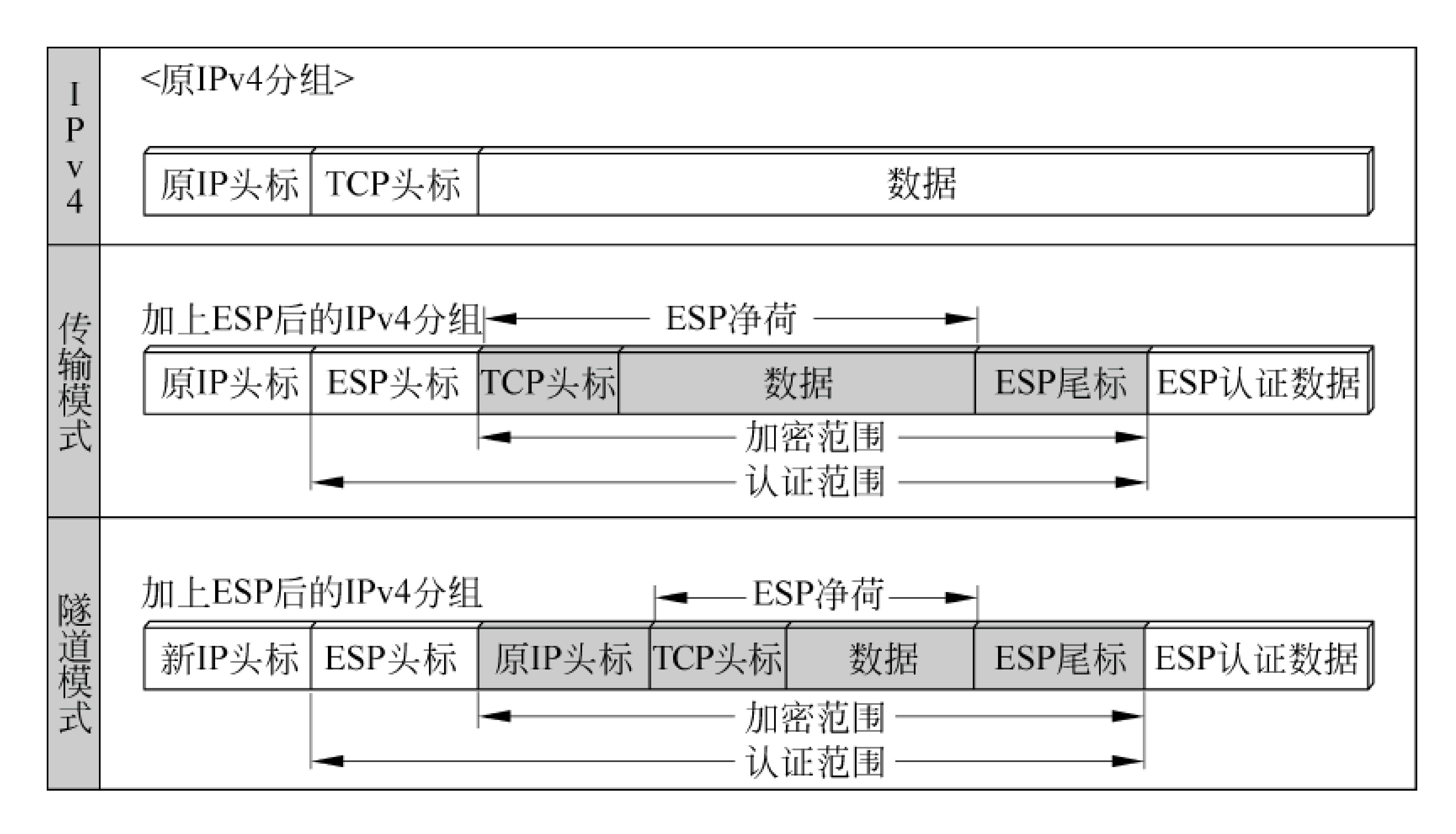


图 2-6　IPv4 中 ESP 的传输模式与隧道模式

1) 传输模式

传输模式仅适用于主机实现，而且仅为上层协议提供保护，而不包括 IP 头。在传输模式中，ESP 插在 IP 头之后和上层协议(例如 TCP、UDP 等)之前，或任意其他已经插入的 IPsec 头之前。

2) 隧道模式

隧道模式可适用于主机和安全网关。整个 IP 数据项被封装在 ESP 有效载荷中，并产

生一个新的 IP 头附着在 ESP 头之前。隧道模式的 ESP 保护整个内部 IP 包,包括源 IP 头。

3. ESP 处理

用 ESP 处理 IP 报文与 ESP 采用的具体模式有一定联系。但无论采用哪种模式,对 ESP 来说,密文是得到验证的,验证的明文则是未加密的,也就是对于外出的包首先进行的是加密处理,而对于进入的包来说首先进行的是验证。使用这种处理顺序能够简化检测过程,抵抗重放攻击以及减少拒绝服务攻击的影响。

1) 外出包处理

在收到上层来的数据包时,IPSec 根据“选择符”(IP 头中的源地址、名字、协议等构成)查询 SPD,寻求匹配条目,决定提供什么服务,随后查询 SADB 寻找合适的 SA。

对在 IPv4 上运行的传输模式应用来说,ESP 头紧跟在 IP 头(包括 IP 头可能有的任何选项)之后,插入一个外出的 IP 包中。IP 头的协议字段被复制到 ESP 头的下一个头字段中,ESP 头的其余字段则被填满:SPI 字段分配到的是来自 SADB 的、用来对这个包进行处理的特定 SA 的 SPI;填充序列号字段的是序列中的下一个值;填充数据会被插入,其值被分配;同时分配的还有填充长度值。随后,IP 头的协议字段得到的是 ESP 的值,或者 50。

除了头插入位置不同之外,IPv6 处理规则基本上类似于 IPv4。ESP 头可插在任意一个扩展头(在路由过程中有可能被修改)之后。

对隧道模式应用来说,ESP 头是加在 IP 包前面的。如果封装的是一个 IPv4 包,那么 ESP 头的下一个头字段分配值为 4;如果封装的是一个 IPv6 包,则分配值 41。其他字段的填充方式和在传输模式中一样。随后,在 ESP 头的前面新增了一个 IP 头,并对相应的字段进行填充(赋值):源地址对应于应用 ESP 的那个设备本身;目标地址取自于用来应用 ESP 的 SA;协议设为 50;其他字段的值则参照本地的 IP 处理加以填充。

两种模式接下来的步骤是相同的,具体如下:

(1) 安全关联查询:得到处理包的策略和 SA,其中包括 SPI、密钥等。

(2) 包加密:在增加了必要的填充项后,使用密钥、加密算法、由 SA 指定的算法模式以及密码同步对载荷数据、填充项、填充长度、下一个头进行加密。如果选择认证,则在认证前要进行加密,加密不包含认证数据字段。由于认证数据不被加密保护,因此要使用认证算法计算 ICV。

(3) 序列号产生:当创建一个 SA 时,发送者的计数器初始化为 0。利用这个 SA,发送的第一个包的序列号设置为 1,计数器的值从此开始递增,并将新值插入到序列号字段,这样就可以保证序列号的唯一性、非零性和单向递增性。

(4) 完整性校验值(ICV)计算:计算 ICV 的参数包含 SPI、序列号、载荷数据(包括初始化向量,原 IP 头、TCP 头和原载荷数据)、填充项、填充长度和下一个头字段的密文数据。

(5) 分段:在进行 ESP 处理后 IPSec 要进行 IP 分段。传输模式 ESP 只适用于整个 IP 数据包,由路由器对 IP 包进行分段,在 ESP 处理之前由接收端进行分段重组。在隧道模式中,应用 ESP 协议处理 IP 包,载荷是分段的 IP 包。

(6) 重新计算位于 ESP 前面的 IP 头校验和,按 IPSec 格式重新封装数据包。

2) 进入包处理

(1) 重组:在 ESP 处理之前执行分段包的重组。

(2) SA 查询:接收到包含 ESP 头的包时,接收者根据目的地址、安全协议和 SPI 查询

单向的 SA。SA 指示出是否检查序列号字段，认证数据字段是否出现，说明解密和 ICV 计算使用的算法和密钥等。

(3) 序列号验证：验证每个接收的包是否包含不重复的序列号。通过使用滑动接收窗口可以拒绝重复序列号。序列号未重复，接收者就进行 ICV 验证。如果 ICV 验证失败，接收者丢弃无效的 IP 数据包，如果 ICV 验证成功，则刷新接收窗口。

(4) ICV 验证：接收者使用认证算法，根据包的字段计算 ICV，验证包的认证数据字段内的 ICV 是否相同。

(5) 包解密：接收者使用密钥、加密算法、算法模式和密码同步数据，对 ESP 载荷数据、填充项、填充长度和下一个头进行解密。在解密数据传输到上一层之前，接收者应检查填充项字段。原始数据包的重组取决于 ESP 的工作模式。

如果篡改了 SPI、目的地址或 IPSec 协议类型字段，那么所选择的 SA 就是不正确的。如果将包映射到另一个这样的 SA，造成的错误和坏包将很难区分。通过使用认证算法，可以检测出 IPSec 头是否已被篡改。如果篡改了 IP 目的地址或 IPSec 协议类型字段，就会发生 SA 不匹配的事情。

2.3.5　IKE 协议

IKE(Internet Key Exchange，Internet 密钥交换)用于动态建立 SA，代表 IPSec 对 SA 进行协商，并对 SADB 进行填充。IKE 建立的基础是 ISAKMP(Internet 安全联盟和密钥管理协议)和两种密钥交换协议(OAKLEY 和 SKEME)。ISAKMP 定义了协商安全的两个独立阶段，阶段 1 是建立通信各方之间已通过的身份验证和安全保护通道，即建立 ISAKMP 的 SA；阶段 2 是交换可为其他协议(例如 IPSec)建立 SA。IKE 利用 ISAKMP 定义密钥交换，进行安全服务的协商，其最终结果是建立 SA。它使用 UDP 协议和 500 端口。

1. IKE 的安全机制

IKE 具有一套自保护机制，可以在不安全的网络上安全地认证身份、分发密钥、建立 IPSec SA。

1) 数据认证

数据认证有如下两方面的概念：

(1) 身份认证：身份认证确认通信双方的身份。支持两种认证方法：预共享密钥(Pre-Shared-Key)认证和基于 PKI(Public Key Infrastructure，公钥基础设施)的数字签名(Rsa-Signature)认证。

(2) 身份保护：身份数据在密钥产生之后加密传输，实现了对身份数据的保护。

2) DH

DH(Diffie-Hellman，交换及密钥分发)算法是一种公共密钥算法。通信双方可以在不传输密钥的情况下通过交换部分数据计算出共享的密钥。即使攻击者截获了双方用于计算密钥的所有交换数据，由于其复杂度很高，不足以计算出真正的密钥。所以，DH 交换技术可以确保双方能够安全地获得公有信息。

3) PFS

PFS(Perfect Forward Secrecy，完善的前向安全性)这种安全特性是指一个密钥被破

解,并不影响其他密钥的安全性,因为这些密钥之间没有派生关系。对于 IPSec,是通过在 IKE 阶段 2 协商中增加一次密钥交换来实现的。PFS 特性由 DH 算法保障。

2. IKE 的交换过程

IKE 使用两个阶段为 IPsec 进行密钥协商并建立 SA。

(1) 第一阶段:通信各方彼此间建立了一个已通过身份认证和安全保护的通道,即建立一个 ISAKMP SA。第一阶段可分为主模式(Main Mode)和野蛮模式(Aggressive Mode)两种 IKE 交换方法。

(2) 第二阶段:用在第一阶段建立的安全隧道为 IPSec 协商安全服务,即为 IPSec 协商具体的 SA,建立用于最终的 IP 数据安全传输的 IPSec SA。

主模式和野蛮模式的相同之处在于都是为了建立保密和验证无误的通信信道,为双方的 IKE 通信提供机密性、消息完整性以及消息源的验证服务。其主要区别在于主模式的消息交换是在一些初始认证完成之后进行的,而野蛮模式则缺少这一层保护,而且交换的消息较少。根据使用的验证方法的不同,主模式和野蛮模式均可以使用 4 种方式实施,分别是预共享密钥认证、数字签名认证、标准公钥加密认证和改正的公钥加密认证,且每一种方式都是使用 Diffie-Hellman 加密钥分配协议来产生密钥因子的。

2.4 传输协议的安全

在传输层上实现数据安全性最早的方法是安全套接层(SSL),另一种是传输层安全协议(TLS)的 Internet 标准,可以为更高层协议(例如 HTTP、FTP、Telnet 等)提供安全服务。一般来说,SSL 或 TLS 可以作为潜在的协议对应用透明,也可以在特定包中使用。

2.4.1 SSL 协议

安全套接层(Secure Socket Layer,SSL)协议主要是使用公开密钥体制和 X.509 数字证书技术保护信息传输的机密性和完整性,它不能保证信息的不可抵赖性,主要适用于点对点之间的信息传输。SSL 是网景(Netscape)公司提出的基于 Web 应用的安全协议,它包括服务器认证、客户认证(可选)、SSL 链路上的数据完整性和 SSL 链路上的数据保密性。对于电子商务应用来说,使用 SSL 可保证信息的真实性、完整性和保密性。但由于 SSL 不对应用层的消息进行数字签名,因此不能提供交易的不可否认性,这是 SSL 在电子商务中使用的最大不足。

SSL 主要考虑在 Internet 和其他 TCP/IP 网络上通信的以下方面的安全保护:

(1) SSL 服务器认证。SSL 服务器认证允许用户确认服务器的身份,支持 SSL 的客户端软件使用标准的公钥密码技术检查服务器的证书和公共 ID 是否有效,并且由用于客户端的可信证书授权(CA)列表中的 CA 颁发证书。

(2) SSL 客户端认证。SSL 客户端认证允许服务器确认用户的身份。采用与服务器认证同样的技术,支持 SSL 的服务器端软件检查客户证书和公共 ID 是否有效,并且由属于服务器端的可信 CA 列表中的 CA 颁发证书。

(3) 加密 SSL 连接。加密 SSL 连接要求所有在客户端和服务器之间发送的信息都被

发送方软件加密，并且由接收方软件解密，以提供高度的机密性。

SSL V3.0 协议由两层组成：SSL 记录协议(SSL Record Protocol)和 SSL 握手协议(SSL Handshake)，如图 2-7 所示。SSL 握手协议允许通信双方在应用协议传输数据之前相互验证、协商加密算法、生成密钥等。记录层封装各种高层协议，具体实施压缩/解压缩、加解密、计算/验证 MAC 等与安全有关的操作。

SSL 握手协议	SSL 修改密文协议	SSL 告警协议	HTTP
SSL 记录协议			
TCP			
IP			

图 2-7　SSL 体系结构

SSL 中有两个重要概念：SSL 连接和 SSL 会话。

(1) SSL 连接：连接时提供恰当类型服务的传输。SSL 连接是点对点的关系，每一个连接与一个会话相关。

(2) SSL 会话：SSL 会话是指客户机和服务器之间的关联，会话通过握手协议来创建。对于每个连接，可以用会话来避免为每个连接进行昂贵的新安全参数的协商。

1. SSL 握手协议

SSL 握手协议使得服务器和客户端能相互鉴别对方的身份、协商加密和 MAC 算法以及用来保护在 SSL 记录中发送数据的加密密钥。在传输任何应用数据前，都必须使用握手协议。

如图 2-8 所示，SSL 握手协议的详细步骤如下：

(1) 客户端(Client)向服务器(Server)发起握手信息 Client Hello，该信息里面包含客户端所支持的所有算法列表和一个用于产生密钥的随机序列。

(2) 服务器收到客户端发来的 Client Hello 信息之后也必须回送一个 Server Hello 信息，Server Hello 信息包含服务器根据客户端的算法列表所选择的一个加密算法、压缩算法和用于密钥建立的随机序列。

(3) 如果需要对服务器进行验证，服务器还需再向客户端发送一个服务器的证书，其中包含服务器的公钥，用于客户端验证服务器端的身份。

(4) 如果没有额外的其他复杂握手信息需要发送，那么服务器向客户端发送一个初始握手完成信息 Server Hello Done。

(5) 客户端和服务器完成以上初始握手信息后，就开始进入到密钥建立阶段。首先，客户端根据它所收到的服务器证书信息来验证服务器的真实身份。如果验证通过，那么客户端提取

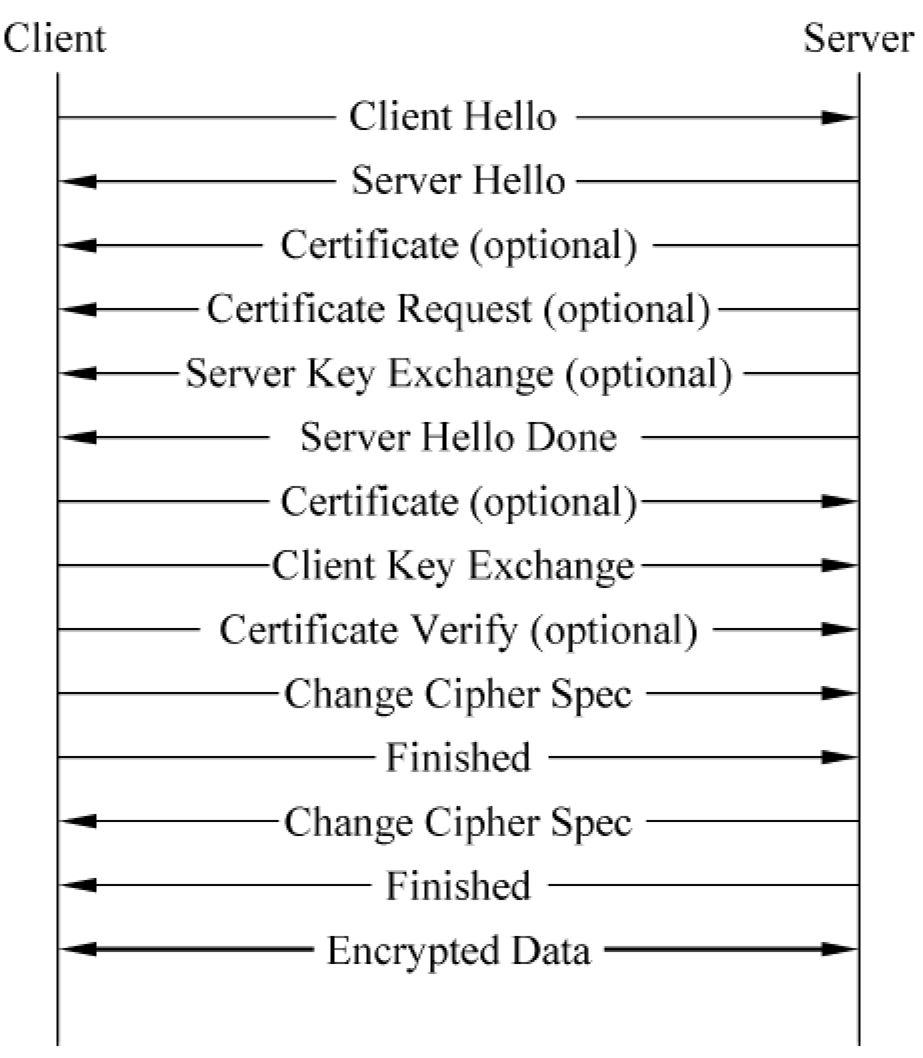


图 2-8　SSL 握手协议详细步骤

证书中的服务器公开密钥,加密一个随机产生的密钥。服务器收到该消息后可以用自己的私钥解密得到该密钥,以后的通信都由该密钥保护。

(6) 客户端向服务器发送更改密码说明消息(Change Cipher Spec),指出用刚刚协商的密钥保护后继会话内容。

(7) 为了防止握手过程的消息被篡改,它包含了对整个连接过程的校验,这样服务器就能判断要使用的加密算法是安全协商的。所以,这步的消息是前面所有消息的MAC值。

一旦服务器收到客户端传来的 Finished 消息,自己也就立刻发送 Change Cipher Spec 和 Finished 消息,表明握手完成,接下来可以进行已加密的数据传输。

2. SSL 记录协议

SSL 协议的底层是记录协议层。SSL 记录协议在客户机和服务器之间传输应用数据和 SSL 控制数据,其间有可能对数据进行分段或者把多个高层协议数据组合成单个数据单元。对记录协议层而言,要封装的高层协议有 4 类:握手协议;修改密文协议;告警协议;应用层协议,例如 HTTP、Telnet 等。

SSL 记录协议为 SSL 连接提供两种服务:

(1) 机密性:握手协议定义了共享的、可以用于对 SSL 有效载荷进行常规加密的密钥。

(2) 报文完整性:握手协议定义了共享的、可以用来形成报文的鉴别码 MAC 的密钥。

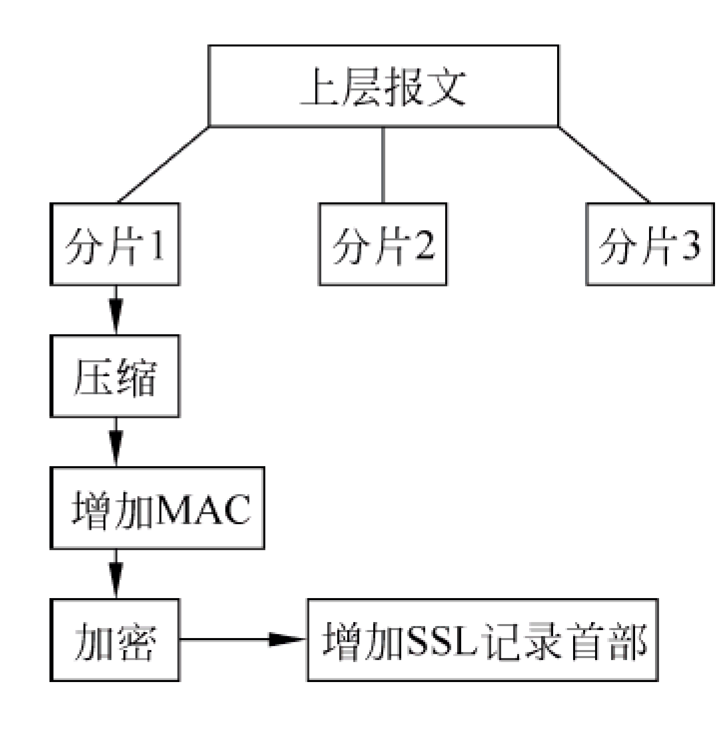


图 2-9 SSL 记录协议的操作

SSL 协议的发送方的具体工作过程如图 2-9 所示,说明如下:

(1) 从上层接收要发送的信息。

(2) 将信息分片。将信息分片为不超过 2^{14}B 的明文记录(Plaintext Records)。

(3) 用当前会话状态指定的压缩算法压缩数据。压缩数据必须做到不丢失信息,并且增加的内容长度不能超过 1024B。

(4) 用当前 Cipher Spec 中指定的 MAC 算法生成 MAC。

(5) 用当前 Cipher Spec 中指定的加密算法加密数据。加密的内容包括压缩报文加上 MAC,加密对内容长度的增加不能超过 1024B。

(6) 在加密数据上附加一个首部。该首部的组成如下:

① 内容类型(8bit):表明用来处理这个包装的数据片的更高层协议。

② 主要版本(8bit):指示使用 SSL 的主要版本。对于 SSLv3 字段值为 3。

③ 次要版本(8bit):指明使用的次要版本。对于 SSLv3 字段值为 0。

④ 压缩长度(16bit):明文数据片以字节为单位的长度(如果使用压缩就是压缩数据片),其最大的值为($2^{14}+2048$)B。

(7) 发送数据。

接收方接收到数据后,处理过程是发送过程的简单逆过程,应该将解密和认证功能倒过来执行。

2.4.2　TLS 协议

TLS(Transport Layer Security)协议定义在 RFC 2246 中,是由 SSL 协议发展而来。TLS 协议与 SSL 协议之间的差别非常小,但是二者不能相互操作。TLS 协议的主要目的也是为了提供通信的机密性和数据的完整性,防止窃听、假冒和信息伪造的威胁。

TLS 协议分为两层：TLS 记录协议(TLS Record Protocol)和 TLS 握手协议(TLS Handshake)。其中 TLS 记录协议是下层协议,它运行在某些可靠传输协议之上,例如 TCP 协议；握手协议层包括握手协议、改变密码规格协议、警告协议等。TLS 的体系结构如图 2-10 所示。

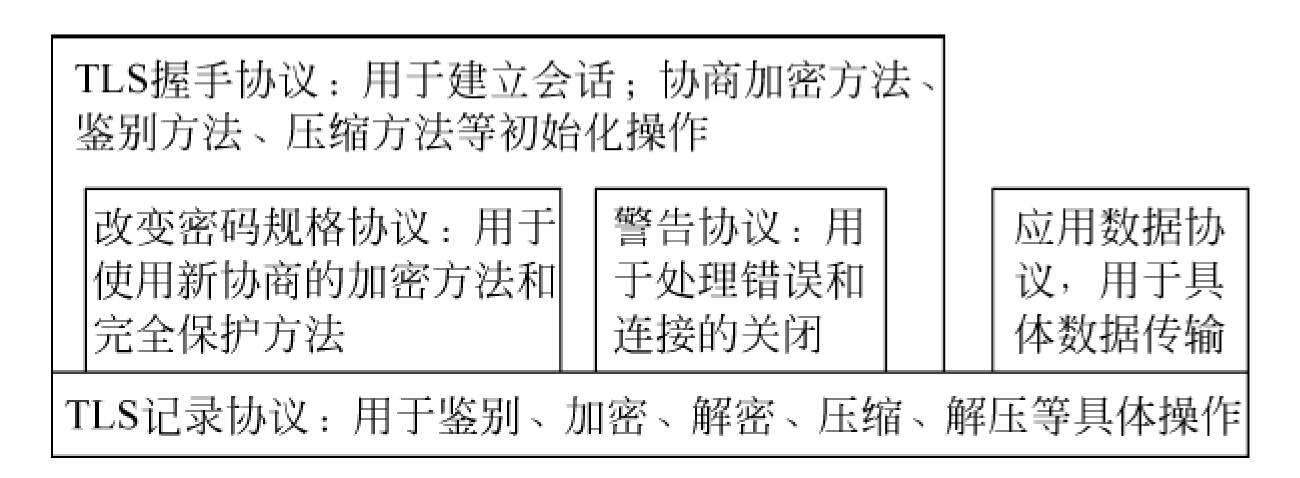


图 2-10　TLS 的体系结构

TLS 利用两种基本方式提供的安全的连接：

(1) 连接的机密性：使用对称密码机制进行数据加密,例如 DES、RC4 等。对称密钥的生成通过秘密协商,每次连接使用不同的密钥。

(2) 连接的可靠性：使用消息认证码来保证消息的完整性,信息鉴别码(MAC)的计算采用 Hash 函数,例如 SHA、MD5 等。

TLS 记录协议用于封装不同的高层协议,TLS 握手协议是其中之一。TLS 握手协议允许服务器和客户相互鉴别,对应用层传输的数据提供加密。TLS 握手协议提供的安全连接有以下 3 个特点：

(1) 使用对称密钥加密算法或公开密钥加密算法(例如 RSA、DSS)来鉴别对等实体的身份,鉴别的方式是可选的,但是必须至少有一方要鉴别另一方的身份。

(2) 协商共享安全信息的方法是安全的,协商的秘密不能够被窃听,而且即使攻击者能够接触连接的路径,也不能获取任何有关连接鉴别的秘密。

(3) 协商是可靠的,没有攻击者能够在不被双方察觉的情况下修改通信信息。

TLS 协议的目标有：

(1) 数据安全：TLS 协议能够被用于在两方之间建立安全连接。

(2) 互操作性：不依赖于应用程序的开发,TLS 协议的一方可以在不知道另一方代码的情况下成功地交换加密信息。

(3) 可扩展性：TLS 协议试图提供一种框架可以使用新的公开密钥和混合密钥(Bulk Encryption Method)方法进行交互。这个目标有两个子目标：避免重新实现一种全新的协议；避免重新实现一个全新的安全库。

(4) 实现高效：加密操作非常耗费 CPU 资源,因此 TLS 协议使用了缓存机制以减少需要建立的连接数。

2.5 应用协议的安全

直接将安全服务嵌入到应用程序中，可以实现特定应用层的通信安全。实现应用层安全的典型例子有安全电子交易协议（SET）、超文本传输协议（HTTP）、远程登入协议（Telnet），下面将对其进行简要介绍。

2.5.1 SET

安全电子交易协议（Secure Electronic Transaction，SET）是由 Visa 和 MasterCard 两大电信公司在 1996 年 2 月发起的，于 1997 年 6 月完成的规范。SET 协议是一个在开放的 Internet 上实现安全电子交易的国际协议和标准，以保证支付信息的机密性、支付过程的完整性、商家及持卡人的合法身份及可操作性。SET 具有以下特点：

（1）信息的保密性。持卡人的账户信息及支付信息在网络传输中是安全的。

（2）数据的完整性。SET 通过引入 RSA 算法的数字签名及 Hash 函数确保这些消息的内容在传输过程中不被非法更改。

（3）持卡人账户认证。持卡人认证是商家提供的验证持卡人是否为合法用户的方式。SET 采用 X.509v3 数字证书和 RSA 数字签名算法来实现这一功能。

（4）商家认证。SET 使持卡人可以鉴别商家的真实性，而且可以验证商家与金融授权是建立了业务联系的，使得商家可以接收信用卡支付。SET 同样采用 X.509v3 数字证书和 RSA 数字签名实现这一功能。

（5）互操作性。互操作性允许在来自不同厂商的硬件和软件使用该规范，并允许持卡人和其他参与者使用它们。

1. SET 系统结构

利用 SET 协议进行的电子商务活动，需要涉及以下 6 个主体：

1）持卡人

进行电子商务活动时一般不涉及现金的交易，大多是通过信用卡来付账，这里将持有发卡行授权使用支付卡的消费者称为持卡人。持卡人应该到发卡银行去申请一套可以进行电子商务活动的软件；此外，持卡人还必须向认证中心 CA 申请一份数字证书，这个证书具有唯一性和权威性。有了数字证书和信用卡就可以进行网上交易等电子商务活动了。

2）商家

有顾客就必然有商家。商家是具有货物或服务卖给持卡人的个人或组织。商家也必须到认证中心申请一份数字证书（可以看成是商家的营业执照）。有了这份数字证书，消费者就能鉴定一个网上商店是否正规。

3）发卡银行

发卡银行是指向顾客提供信用卡的金融机构。顾客购买商品，进行网上支付的时候，扣款必须经过发卡银行授权。发卡银行并不是 SET 交易的直接组成部分，但它却是完成交易必要的参与方。一般来说，顾客向认证中心申请数字证书也必须得到发卡银行的批准。

4）收单银行

同发卡银行一样，收单银行是向商家提供账户的金融机构。顾客支付给商家的数字货币也是直接划在商家在收单银行的账户上。收单银行也不是 SET 交易的直接组成部分，但它是完成交易的必要参与方。

5）支付网关

SET 交易必须有发卡银行、收单银行的参与。但是，银行的主机并不能直接连接在 Internet 上。为了能够接收从网上传来的付账信息，必须有一个专用系统完成这种中转，这个专用系统就是支付网关。支付网关接收从 Internet 上传来的支付信息，然后将支付信息发给收单银行，收单银行和发卡银行之间进行银行的内部转账，完成支付。

6）认证中心

认证中心可以说是 SET 交易中最重要的一员。如果没有认证中心的参与，SET 交易是不能完成的（商家、顾客、支付网关都必须持有认证中心颁发的数字证书）。认证中心的主要功能包括证书的签发、存储、目录服务、证书的挂失和更新以及密钥的更新及恢复等。

2. SET 的安全机制

SET 协议综合利用了密码学的各项技术来保证交易信息的安全。

（1）数据发送者可以随机生成对称密钥。用对称密钥加密数据，并将对称密钥用接收方的公开密钥加密装入数字信封，此数字信封只能用接收方的私人密钥解密打开，可以保证数据的安全和保密。

（2）通过消息摘要的检验，可以保证数据的完整性。消息发送方将要发送的消息经哈希运算，计算出此条消息的消息摘要，消息接收方在收到消息的同时也收到此消息摘要，接收者再用同样算法计算出他所收到的消息摘要，并将计算出的消息摘要与收到的消息摘要进行比较，如果两条消息摘要相同，说明消息在传输过程中未被篡改。

（3）使用经 CA 签名的数字证书，可以认定双方的身份。由于证书是由大家公认的权威机构 CA 在对用户进行认证后发给用户的，并且 CA 还在证书上用自己的私人密钥对所发证书的消息摘要进行加密，生成 CA 的数字签名，消息接收方在收到对方的数字证书后，可以用 CA 的公开密钥对 CA 的数字签名解密。证明对方发来的证书确实是 CA 发的，也就可以证明对方的身份。

（4）通过验证对方的数字签名可以确认消息确实是对方发的，从而保证了交易的不可抵赖性。因为数字签名是用消息发送方的私人密钥对所发消息的消息摘要进行加密生成的，只要能用消息发送方的公开密钥解密，就能证明此数字签名肯定是消息发送方生成的，又因为数字签名所加密的消息摘要是由所发消息生成的，可以证明消息确实是由消息发送方发送的。

（5）在 SET 交易中，持卡人要发给网关的信息是通过商家转交的。持卡人用网关的公开密钥对持卡人要发给网关的信息进行加密，并将信息装入数字信封内（只能由网关打开），同时采用双重签名的方法，使商家不能看到持卡人发给网关的信息，网关也不能看到持卡人发给商家的信息。

3. SET 的工作过程

1）发送方的处理

发送方在发送前必须先取得两个证书：一个是自己的证书；另一个是接收方的证书。

只有具备了两个证书才能正式进行通信。整个通信流程如下：

(1) 发送方对接收方的数字证书进行认证。如果验证通过，继续执行以下步骤；否则中止协议。

(2) 发送方对要发送的消息明文进行哈希运算，生成消息摘要，以便于进行数字签名。

(3) 发送方利用自己的私人密钥对消息摘要加密，生成数字签名。

(4) 随机生成一个对称密钥用于加密消息。

(5) 用对称密钥对消息明文加密，生成消息密文。

(6) 用接收方的公开密钥对对称密钥加密，装入数字信封。

(7) 将消息密文、数字签名、数字信封及数字证书一起发给接收方。

2) 接收方的处理

接收方按照以下步骤对所接收的消息进行解密和检验。

(1) 接收方首先对发送方发来的数字证书进行签名认证。如果认证通过，协议继续执行，否则中止协议。

(2) 认证通过后，接收方利用自己的私钥对数字信封解密，得到用于加密信息的对称密钥。

(3) 用对称密钥对消息密文解密，得到消息明文。

(4) 用发送方的公开密钥对数字签名解密，得到消息摘要。这其实也是一个签名验证。

(5) 接收方对消息明文进行哈希运算，得到新的消息摘要。

(6) 比较两个消息摘要，确认消息的完整性。

(7) 保存消息明文。

2.5.2 HTTP

超文本传输协议(Hypertext Transfer Protocol，HTTP)是万维网(WWW)服务器与浏览器之间信息传输规范的网络协议，是目前 Internet 上使用最广泛的应用层协议。HTTP 协议运行在客户程序和服务器程序中，不同端系统上的客户程序和服务器程序通过交换 HTTP 消息彼此交流。

HTTP 的工作过程为典型的客户机/服务器工作模式，客户即为 WWW 浏览器，服务器即是 WWW 服务器。如图 2-11 所示，HTTP 的工作过程使用请求/响应握手方式：(1)客户机与服务器发生连接，双方建立起一个 TCP 连接；(2)客户机发出请求；(3)服务器处理请求，返回应答；(4)服务器关闭连接。

1. HTTP 的信息传输方式

HTTP 具有两种信息传输方式：

1) 点对点方式

这显然是最简单的传输方式，拨号上网用户访问自己接入的 ISP(Internet 服务提供商)的 WWW 服务器即为此种方式。客户端主机与源服务器之间建立起点对点的直接连接关系，响应速度当然也是比较快的。

2) 借助中间服务的方式

该方式也称为设定所谓中间服务器，或称中

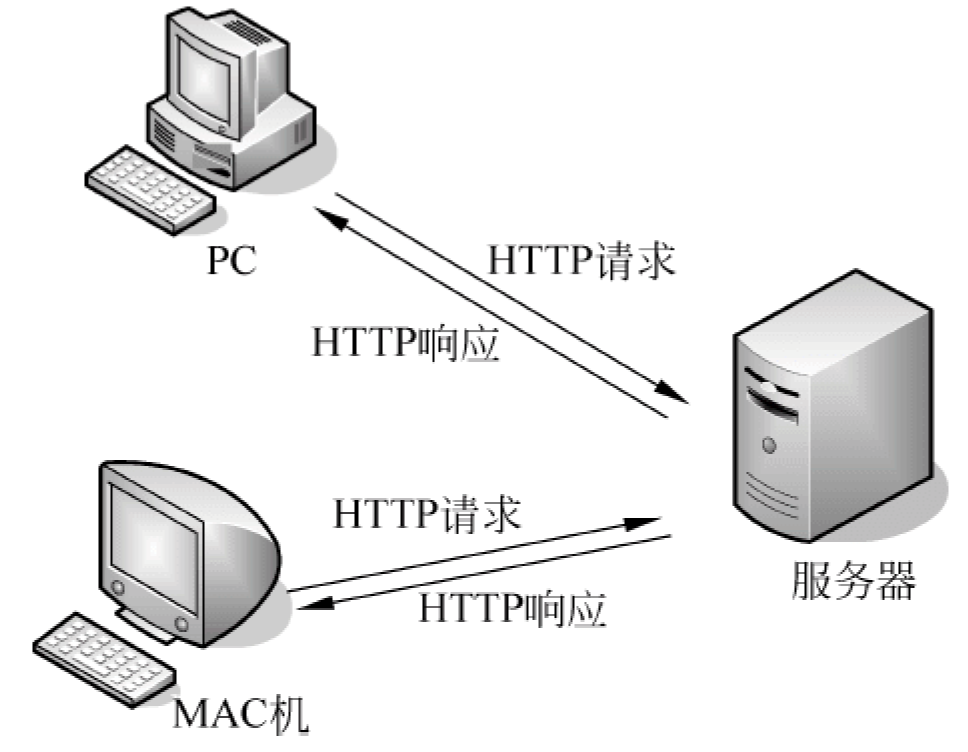


图 2-11 HTTP 的工作模式

继节点。HTTP 支持 3 种类型的中间系统。

(1) 代理服务器(Proxy)：接收客户发送的 HTTP 资源请求，代替客户向服务器发送请求，然后代替客户接收服务器的应答，再转给客户。从功能上看其具有双重性，既作为客户向源服务器发请求，又可作为服务器向客户返回应答。

(2) 网关型中间服务器(Gateway)：它可以代替那些不能够直接与客户通信的源服务器接收客户请求，对于客户来说，Gateway 就是源服务器，可用作防火墙。

(3) 隧道型服务器(Tunnel)：它对于客户和源端服务器来说，都是不可见的，它只是简单地把接收到的 HTTP 数据流转发出去，对数据流本身不作任何改变，也就是简单的中继作用。除隧道服务器外，中间服务器、源服务器系统都可具有本地的缓存机制。

2. HTTP 的消息结构

为在客户浏览器与服务器之间交换信息，完成传输过程，HTTP 定义了一套完整的消息结构。消息分两类：客户方发出的请求消息；服务器发出的响应消息。这两者都依据制定传输类体(亦即消息载体)的 RFC 822 规范中规定的通用消息格式具体实施。

1) 请求消息

请求消息(Request Message)：由客户端发给服务器的消息。其组成包括请求行(Request-Line)、可选的头域(Header Field)及实体(Entity-Body)。

请求消息结构：

```
Full - Request  =  Request - Line
                    * (General - Header
                        | Request - Header
                        | Entity - Header)
                    CRLF
                    [Entity - Body]
```

请求行结构：

```
Request - Line  = Method SP
                  Request - URI
                  SP
                  HTTP - Version CRLF
```

HTTP 协议请求响应方法如表 2-2 所示。

表 2-2　HTTP 协议请求响应方法

方 法 名	备　　注
GET	获取一个 URL 指定的资源，即资源实体
HEAD	获取一个指定资源的信息
POST	向服务器提交数据
PUT	向服务器提交资源
DELETE	请求源服务器删除 Request-URI 标识的资源
TRACE	网络跟踪
CONNECT	与 Proxy 之间的连接管理
OPTIONS	查询能力

2）响应消息

响应消息(Response Message)：是服务器端回复客户端请求的消息，其组成包括状态行(Status-Line)、可选的头域(Header Field)及实体(Entity-Body)。

响应消息结构：

```
Full - Response  = Status - Line
                   * (General - Header
                       | Response - Header
                       | Entity - Header)
                     CRLF
                   [Entity - Body]
```

状态行结构：

```
Status - Line  = HTTP - Version SP
                 Status - Code SP
                 Reason - Phrase CRLF
```

HTTP 协议响应消息及说明如表 2-3 所示。

表 2-3　HTTP 协议响应消息

状态码	定　　义	说　　明
1××	信息	接收到请求，继续处理
2××	成功	操作成功时收到，理解和接受
3××	重定向	为了完成请求，必须采取进一步措施
4××	客户端错误	请求的语法有错误或不能完全被满足
5××	服务器端错误	服务器无法完成明显有效的请求

3. HTTP 的缺陷及策略

HTTP 协议虽然使用极为广泛，但存在不小的安全缺陷，主要体现在数据的明文传输和消息完整性检测的缺乏，而这两点恰好是网络支付、网络交易等新兴应用中安全方面最需要关注的。

针对 HTTP 协议的安全缺陷，HTTPS(Hypertext Transfer Protocol over Secure Socket Layer)通过在 TCP 层和 HTTP 层之间增加 SSL 来加强安全性，数据传输过程中，加/解密均由 SSL 进行，而与上层的 HTTP 无关，对于 HTTP 来说是透明的。HTTPS 增强的安全性主要表现在：

(1) 双向的身份验证。

(2) 数据传输的机密性。

(3) 数据的完整性检验。

(4) 防止数据包的重放攻击。

2.5.3　Telnet

远程登录协议(Telecommunication Network Protocol，Telnet 协议)是 TCP/IP 协议族

中的一员。Telnet 协议能够把本地用户所使用的计算机变成远程主机系统的一个终端。远程登录的思想体现了层次结构概念，远程登录的实现使本地用户并不直接面对远地系统的各种资源，相当于在服务客户与具体服务之间加入一个中间层次，即远程登录服务器。

1. Telnet 协议的主要内容

1）网络虚拟终端（NVT）

网络上进行通信的两台主机所使用的字符集可能不一样，Telnet 协议为通信的两台主机（包括打印机和键盘）提供了一个标准接口——网络虚拟终端（Network Virtual Terminal，NVT）。本地主机和远程主机都必须将自己的终端特性转换为统一的网络虚拟终端，把主机从维护与它通信的终端特点任务中解放出来，从而各自不用了解对方主机的内部细节而直接建立通用的应用程序。

2）选项协商

选项协商是 Telnet 协议最复杂的部分，总共有 39 个选项用于配置本地和远程主机间的工作模式。当一方要执行某个选项时需向另一端发出请求，若对方接受该选项，则选项在两端同时起作用，否则两端保持原来的模式。Telnet 的命令格式如图 2-12、图 2-13 所示。IAC 是 Telnet 协议中的保留码，双方用 IAC 确定收到的字节是数据还是命令。Telnet 协议的命令是至少包含两个字符（IAC 和命令码）的字节序列，选项协商则有 3 个字节，第 3 个字节为协商的选项。当协商的选项存在子选项时，要进行子选项协商，命令码如表 2-4 所示。

IAC	命令码	选项码

图 2-12　Telnet 选项协商命令格式

IAC	SB	选项码	参数	IAC	SE

图 2-13　Telnet 子选项协商命令格式

表 2-4　Telnet 命令码

名　称	值	意　　义	名　称	值	意　　义
SE	240	子选项协商结束	Erase Line	248	删除行
NOP	241	空操作	Go Ahead	249	继续传输符号
Data Mark	242	紧急数据部分	SB	250	开始子选项协商命令
Break	243	BRK 信号	WILL	251	执行协商的选项
Interrupt	244	IP 信号	WONT	252	拒绝执行协商的选项
Abort Output	245	AO 信号	DO	253	请求执行协商的选项
Are You There	246	AYT 信号	DONT	254	要求停止协商的选项
Erase Character	247	删除字符	LAC	255	保留码

2. Telnet 协议的实现原理

Telnet 协议原理如图 2-14 所示：用户的数据或命令通过 NVT 键盘输出，将命令交给 NVT 命令执行体完成相应的功能；字符则由 Socket 经过网络发送给远程系统，远程系统发回的命令和数据由 Socket 交给命令执行体；若是命令则执行相应的功能，字符直接显示在 NVT 打印机（显示器屏幕）上。

NVT 键盘有一个字符缓冲区，默认情况下当缓冲区满时，才发送数据，如果双方协商过程中同意执行 echo 和 Go Ahead 选项，那么 NVT 键盘每产生一个字符就立即发送出去。

本程序通过函数 NVT_CharHToN(char * pChar)将本机键盘产生的键值转换为对应的 NVT 字符,然后发送出去。

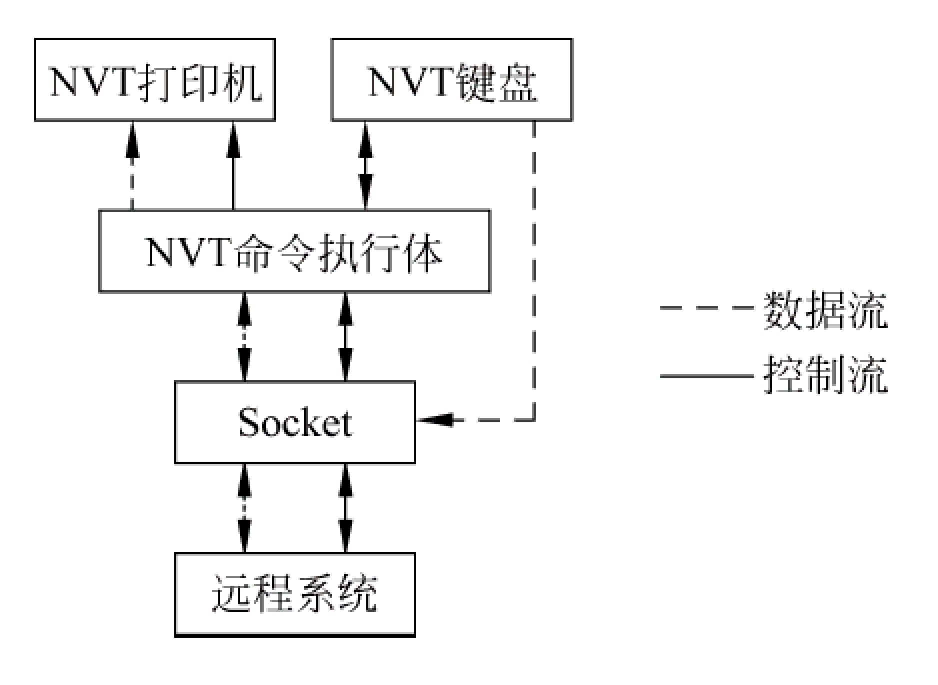


图 2-14 Telnet 协议原理结构图

3. Telnet 的安全性分析及策略

Telnet 使用广泛,应用也很方便,其登录过程需要进行用户的身份认证,表面上是很安全的服务。其实,Telnet 存在着严重的安全隐患。

(1) 所有的数据在传输过程中都没有任何加密措施,所以很容易被攻击者利用网络嗅探器捕获,进而遭受攻击。

(2) Telnet 没有用户的强身份认证措施,攻击者可以对每个账户的密码进行穷举攻击。虽然这些错误猜测将被记录在日志文件中,但 Telnet 本身并不记录猜测的系数。

(3) Telnet 本身不进行会话的完整性检查,而数据全都是明文传输,容易被非法篡改。

针对上述安全隐患,可以采用一些增强 Telnet 协议的安全性措施,下面将对其简要介绍。

1) IP 分组过滤

使用 IP 分组过滤可以增强 Telnet 协议的安全性。IP 分组过滤是在网络层和传输层提供的分组过滤功能,通过对网络层 IP 数据包中的源/目的地址、源/目的端口号以及协议类型 5 类信息的识别来实现。通常,网络管理员应根据事先制定的安全策略确定并配置上述 5 类信息的分组过滤规则,路由器的任何物理端口都可根据需要引用已经确定的若干条过滤规则。分组过滤过程就是检查输入的 IP 数据包符合哪条分组过滤规则,如果符合则转发这个 IP 数据包,否则再用其他过滤规则检查或直接丢弃。

2) 加密隧道技术

加密隧道是指网络接入设备(例如路由器)能自动识别网络上主机发来的信息,有选择地对特定的信息加密后在网上传输,可能经过若干普通路由器后到达目标路由器,信息仅在目标路由器处解密,从而实现对网络上其他无关设备屏蔽的隧道效应。采用加密隧道技术可以把某些特定的 Telnet 连接通路变为加密的隧道。形成隧道的具体步骤是:

(1) 进行路由器身份认证和广播公钥。

(2) 在传输数据前,产生并加密传输消息密钥(双密钥加密技术),并指定相应的加密条件。

(3) 产生本次传输使用的会话密钥,开始进行数据传输(序列密钥加密技术)。

3) 一次性密码认证技术

在 Telnet 的 TCP 连接建立后,要求登录的用户提供自己的账户和密码,在客户端不做修改的情况下,账户和密码至少在本地局域网上是明文传输的。采用一次性的密码认证技术,用户在进行 Telnet 过程中,即使用户账户和密码被窥探也不影响安全性。OTP(One-time Password,动态密码)可以产生一系列一次性密码,用户使用这些密码使自己的身份在远程系统上得到认证,而不必担心被窥探。OTP 中,某个一次性密码使用完后,用户再次登录则需要新的一次性密码,这使得安全性得到有效保障。

4）Telnet 代理服务器

为了允许远程终端通过防火墙(Firewall)实现访问，通过一个 Telnet 代理服务器，它支持基于 IP 地址或主机名的访问控制，也支持允许选择任何目标被阻塞的访问控制，记录所有连接和传输的字节。使用 Telnet 代理服务器不会对防火墙系统本身造成任何威胁，因为它运行时不允许改变到被限制的目录，而且整个 Telnet 代理服务器的菜单处理都在内存，也不需要 shell 或程序。

系统管理员希望能够登录到防火墙系统进行维护和管理，有两个可供选择的方案。

(1) 配置防火墙系统的 Telnet 服务器进程 telnetd，通常的做法是设置 Telnet TCP 服务端口(23)，管理员必须从控制台登录。

(2) 在防火墙系统的 inetd.conf 文件里配置 telnetd，以便对 telnetd 的访问用网络控制文件(netcal)保护。

思　考　题

(1) TCP/IP 协议可分为哪几层，每一层的安全协议有哪些?

(2) 针对网络协议的攻击有哪些?

(3) 请简述 DDoS 攻击原理。

(4) IPSec 安全体系结构中包括了哪几种最基本的协议?

(5) 什么是 AH 协议? AH 可采用哪两种模式? 请指出 AH 在这两种模式中的区别。

(6) 什么是 ESP 协议? ESP 可采用哪两种模式? 请指出 ESP 在这两种模式中的区别。

(7) 在同一端到端通信上既使用 AH 协议又使用 ESP 协议时，应怎样安排 AH 和 ESP 的应用顺序，为什么?

(8) 写出并分析一次完整的 SSL 连接过程。

(9) TLS 协议包含哪些协议? 请简述各个协议的作用。

(10) 简述 SET 协议的工作过程。

(11) 请写出 HTTP 请求消息和响应消息的格式。

(12) 增强 Telnet 协议安全的措施有哪些? 请分别加以简要介绍。

参 考 文 献

[1] 郝玉洁，刘桂松，秦科，等. 信息安全概论. 成都：电子科技大学出版社，2007.

[2] 毛云飞，刘笑凯，郑连清. TCP/IP 协议的安全性分析及对策. 计算机应用研究，2003，(4)：88～90.

[3] 张霞. TCP 协议的安全性分析. 网络与信息，2009，(7)：24～25.

[4] Naganand Doraswamy 著. IPSec 新一代 Internet 安全标准. 京京工作室译. 北京：机械工业出版社，2000.

[5] 龚俭，杨望. 计算机网络安全导论. 南京：东南大学出版社，2007.

[6] 凌捷，谢赞福. 信息安全概论. 广州：华南理工大学出版社，2005.

[7] 梁军，毛振寰. 计算机网络与信息安全. 北京：北京邮电大学出版社，2005.

[8] 秦科,张小松,郝玉洁.网络安全协议.成都:电子科技大学出版社,2008.
[9] 洪帆,崔国华,付小青.信息安全概论.武汉:华中科技大学出版社,2005.
[10] 唐建雄.IPSec体系结构分析及实现策略.武汉:交通与计算机,2001,(2):41~44.
[11] Postel J,Reynolds J. RFC 854:Telnet Protocol Specification. 1983.
[12] 肖戈林.HTTP协议技术探析.江西通信科技,2001,(1):17~24.
[13] 卢爱卿,张会勇,赵征.Telnet协议的实现原理及应用.计算机工程,2002,(28):268~280.
[14] 陈修还,石岩,王栋.增强使用Telnet协议的安全性.计算机工程与应用,1999,(2):90~92.

第3章　信息加密与认证技术

本章学习目标

密码技术是保护数字内容安全的一个重要手段,而信息认证是判断数字内容完整性的重要技术。本章将介绍密码学的基本原理,主要包括古典密码学、对称密码技术以及非对称密码技术,并且介绍信息认证的基本概念、单向 Hash 函数与消息认证码的基本原理、典型的认证方法和认证技术。

通过对本章的学习,应掌握以下内容:

(1) 古典密码技术的分类和基本原理。

(2) 对称密码技术与 DES、AES 算法。

(3) 公钥密码技术与 RSA、ElGamal、ECC。

(4) 信息认证的概念与作用及其基本原理。

(5) 单向 Hash 函数与消息认证码的基本概念和原理。

(6) 数字签名的原理和技术。

(7) 身份认证的典型技术。

信息加密技术是一种利用数学或者物理手段对电子信息在传输过程中和存储介质体内进行保护以防止泄漏的技术。保密通信、计算机密钥、防复制软盘等都属于信息加密技术。它是对付各种安全威胁的最强有力的工具。认证是防止主动攻击的一项重要技术,可以用于开放环境中各种信息系统安全性的保护。

本章将介绍密码学中的一些基础知识和常见的密码学技术,并对信息认证的数字签名技术和身份认证技术进行详细介绍。本章的学习将为后面章节的学习打下基础。

3.1　密码学技术概述

密码学包括密码编码学(Cryptography)和密码分析学(Cryptoanalytics)两部分,这两部分既相互对立又相互促进。密码编码学是一种信息保护技术,主要研究如何编码及采用怎样的编码体制来改变被保护信息的形式,使得加密后的信息除指定接收者之外的其他人都不可理解。与密码编码学相对应的是密码分析学,密码分析学是一种破译密文的技术,主要研究在未知密钥的情况下从密文中推导出明文或密钥的技术。

3.1.1　密码系统的组成

密码系统是用于对消息进行加密、解密的系统。可以用一个五元组来表示密码系统,即明文、密文、密钥、加密算法和解密算法。

(1) 明文(Plaintext):未加密的原始信息(数据)。

(2) 密文(Ciphertext):明文经变换后,即被加密后的结果。

(3) 密钥(Key):参与密码变换的参数。

(4) 加密(Encryption):将明文变换为密文的过程。

(5) 解密(Decryption):加密的逆过程,即由密文恢复出明文的过程。

(6) 加密算法(Encryption Algorithm):密码员对明文进行加密时所采用的一组规则。

(7) 解密算法(Decryption Algorithm):接收者对密文进行解密时所采用的一组规则。

一个密码或者密码体制用于加密数据的过程如下:在发送端,原始数据(明文)通过加密形成密文,在接收端通过解密将密文恢复成明文,在密码体制中加密和解密要用到的密钥分别是加密密钥和解密密钥。传统密码体制模型如图 3-1 所示。

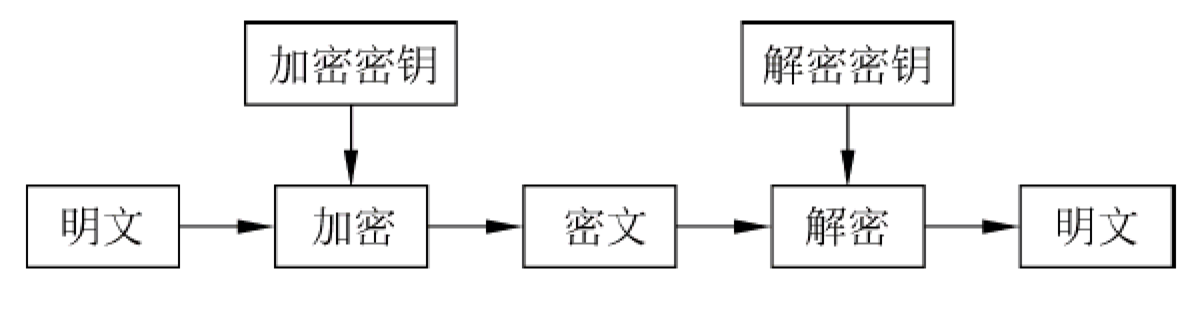


图 3-1 传统密码体制模型

3.1.2 密码学的分类

数据加密算法有很多种,密码算法的标准化是现代信息化社会发展的一个必然趋势。按照不同的标准,密码学的分类也有所不同,下面介绍几种常用的分类方法。

1. 古典密码学和现代密码学

按照密码学的历史发展阶段划分,密码学可以分为古典密码学和计算机密码学两个阶段。

1) 古典密码学

古典密码学是密码学发展的第一个阶段,又称为传统密码学阶段。古典密码学主要依靠人工和机械进行信息的加密、传输和破译。古典密码学加密的对象是文字信息,其内容都是基于字母表(例如英文字母表、汉语拼音字母表等)。古典密码系统的加密算法主要有替代加密、置换加密等。

2) 现代密码学

这是密码学发展的第二个阶段,亦称为计算机密码学阶段。现代密码学利用计算机进行自动或半自动的加密、解密和传输。计算机密码学加密的对象是计算机系统所使用的数据,也就是普遍采用的二进制数据。以二进制的数字化信息为研究对象,并使用现代思想进行信息的保密,这是现代密码学的一个显著特点。现代密码学发展至今,根据密钥的使用方式又可分为对称密钥密码和非对称密钥密码两个发展方向。

2. 对称密钥密码和非对称密钥密码

1) 对称密钥密码(Symmetric Cryptography)

不管是在加密还是解密的过程,都需要有密钥的参与。如果用于加密数据的密钥和解密数据的密钥相同或者二者之间存在着某种明确的数学关系(即很容易由其中一个密钥推导出另外一个密钥),这样的密码体制就称为对称密钥密码体制。对称密钥密码体制又称为私密钥密码体制,它的加密密钥和解密密钥都是要保密的。

由于对称密钥密码体制所使用的加密密钥和解密密钥相同,也称为单钥密码体制。该

类型密钥密码体制的主要算法有 DES、IDEA、TDEA、MD5、RC4 和 AES 等。

2）非对称密钥密码（Asymmetric Cryptography）

如果用于加密数据的密钥与用于解密数据的密钥不相同，而且从加密的密钥无法推导出解密的密钥，这样的密码体制就称为非对称密钥密码体制。非对称密钥密码体制中，往往其中一个密钥是公开的，另一个是保密的。

由于非对称密钥密码体制中有一个密钥是可以公开的，所以又可称为公开密钥密码体制。非对称密钥密码体制的主要算法有 RSA、E1gamal、Rabin、DH 和椭圆曲线等。

3. 分组密码和序列密码

按明文加密时的处理方法划分，密码体制可以分为分组密码（Block Cipher）体制和序列密码（Stream Cipher）体制两种。

1）分组密码（Block Cipher）

如果密文仅与给定的密码算法和密钥有关，与被处理的明文数据段在整个明文（或密文）中所处的位置无关，则这种密码体制就叫做分组密码体制。

分组密码加密时，首先将明文序列以固定长度（例如 32bit）进行分组，每组明文用相同的密钥和算法进行变换，得到一组密文。分组密码是以块为单位，在密钥的控制下进行一系列线性和非线性变换而得到密文的。加密算法中重复地使用替代和移位两种基本的加密变换，使用打乱（替代）和扩散（移位）技术对信息进行隐藏。打乱就是改变数据块，使输出位与输入位之间不具备明显的统计关系，而扩散就是通过密钥位转移到密文的其他位置上。

2）序列密码（Stream Cipher）

如果密文不仅仅与给定的密码算法和密钥有关，同时也是被处理的明文数据段在整个明文（或密文）中所处位置的函数，则这样的密码体制就称为序列密码体制，又称为流密码体制。序列密码的加密过程是把报文、语音以及图像等原始信息转换为明文数据序列，再将其与密钥序列进行异或运算，生成密文序列发送给接收者。接收者使用相同的密钥序列与密文序列再进行异或运算，从而恢复出明文序列。

序列密码加密和解密的密钥通常是采用比特流发生器随机产生二进制比特流而得到的，它与明文结合产生密文，与密文相结合可以产生明文。序列密码的安全性主要依赖于随机密钥序列。

3.2　古典密码技术

在计算机出现前，密码学由基于字符的密码算法构成。不同的密码算法是字符之间互相代替或者互相换位，好的密码算法是结合这两种方法重复进行多次运算。现代密码学变得越来越复杂，但是其基本原理仍是一致的，不同之处在于算法是针对比特而不是对字母进行变换，实际上这只是字母表长度上的改变，即从 26 个元素变为 2 个元素。大多数好的密码算法仍然是代替和换位的元素组合。本节将介绍几种典型的古典密码，这是密码学的基础，每一种密码技术都有其独特之处。

3.2.1 代替密码

代替密码是古典密码中常用到的两种基本处理技巧之一，它在现代密码学中依然得到了广泛应用。所谓代替，就是将明文中的字母用其他字母、数字或符号所取代的一种方法。常见的代替密码技术包括单表代替密码和多表代替密码。

1. 单表代替密码

单表代替密码对明文中的所有字母都使用同一映射，即 $\forall p \in P, f: P \rightarrow C, c = f(p)$。为了确保解密的正确性，通常要求映射 f 是一一映射。提到单表代替密码就不得不先介绍凯撒(Caesar)密码。凯撒密码作为一种最为古老的对称加密体制，在古罗马的时候就已经很流行，它的基本思想是：通过把字母移动一定的位数来实现加密和解密。例如，如果密钥是把明文字母的位数向后移动 3 位，则位数就是凯撒密码加密和解密的密钥，这时明文与密文的对应如表 3-1 所示。

表 3-1 凯撒密码明文与密文对照表

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
对应数字	0	1	2	3	4	5	6	7	8	9	10	11	12
密文	D	E	F	G	H	I	J	K	L	M	N	O	P
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
对应数字	13	14	15	16	17	18	19	20	21	22	23	24	25
密文	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

表 3-1 给出的仅为向后移动 3 位的凯撒移位，但显然从 1～26 个位置的移位都可以使用，将凯撒密码通用化就可以得到移位代替密码。

1) 移位代替密码

设：$P = C = K = Z_{26}$，这里，P、C、K、Z_{26} 分别表示明文空间、密文空间、密钥空间和 26 个整数(对应的 26 个英文字母)组成的空间。对于任意大小 $k \in K$，可以得到加密过程如下：

$$E_k(p) = p + k(\text{mod } 26) = c \in C \tag{3-1}$$

其中，p 为明文，c 为密文，k 为密钥。

解密过程如下：

$$D_k(c) = c - k(\text{mod } 26) = p \in P \tag{3-2}$$

移位代替密码算法是不太安全的，由于模为 26，所以只存在 26 个可能的密钥，即需要测试的密钥仅为 25 次。它可被穷举密钥搜索所分析。另外，26 个英文字母在文字信息中的出现有一定的统计规律，单表替代密码算法由于没有把不同字母出现的频率隐藏起来，破译起来比较容易，不能抵抗明文统计特性的攻击。

例 3-1 对于凯撒密码，当 $k=3$ 时，代替表如表 3-1 所示。

若明文为 p=casear cipher is a shift substitution 时，密文为 c=FDVHDU FLSKHU LV D VKLIW VXEVWLWXWLRQ。

解密时只需要用密钥 $k=3$ 的加密密钥对密文 c 进行解密运算就可以恢复出原文。

这种密码是将明文字母表中字母位置下标与密钥 k 进行模 26 加法运算，所得的结果作为密文字母位置下标，相应的字母即为密文字母。

2）乘法代替密码

已知 $p=c=k=z_{26}$，k 是满足 $0<k<n$ 的正整数，要求 k 与 n 互素。

加密算法如下：

$$c=E(k,p)=(pk)(\bmod n) \tag{3-3}$$

解密算法如下：

$$p=D(k,c)=k^{-1}c(\bmod n) \tag{3-4}$$

注意：乘法代替算法要求 k 与 n 互素的原因是仅当 $\gcd(k,n)=1$ 时，才存在两个整数 x、y 使得 $xk+yn=1$，才有 $xk\equiv 1 \bmod n$，进而有 $p\equiv xc \bmod n$，明文和密文才是一一对应的，密码才能正确解密。

例 3-2　英文字母表 $n=26$，$k=9$。则有乘法代替密码的明文与密文字母对应表，如表 3-2 所示。

表 3-2　乘法代替密码明文与密文对照表

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
密文	A	J	S	B	K	T	C	L	U	D	M	V	E
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
密文	N	W	F	O	X	G	P	Y	H	Q	Z	I	R

对照表 3-2，若明文为 p=multiplicative cipher，则其对应的密文为 c=EYVPUFVUSAPUHKSUFLKX。

3）仿射密码

乘法密码和加法密码二者相结合便可构成仿射密码。仿射密码是一种线性变换。对于 $p=c=k=z_{26}$，且 $K=\{(a,b)\in z_{26}\times z_{26},\gcd(a,26)=1\}$，对于任意的 $k=(k_1,k_2)\in K$，加密算法如下：

$$c=E(k,p)=k_1p+k_2(\bmod 26) \tag{3-5}$$

解密算法如下：

$$p=D(k,c)=k_1^{-1}(c-k_2)(\bmod 26) \tag{3-6}$$

其中，式(3-6)中的“-1”表示“逆”。显然，当 $k_1=1$ 时，仿射密码为对应为凯撒密码。仿射密码共有($26\times 12=312$)个可能的密钥，其中 12 是满足 $\gcd(a,26)=1$ 的 a 的个数。

例 3-3　设 $k=(k_1,k_2)=(5,3)$，可以计算得到：$5^{-1}(\bmod 26)=21$；仿射密码的加密函数为 $c=5p+3(\bmod 26)$；相应的解密函数为 $p=21(c-3)(\bmod 26)=21c-11(\bmod 26)$。

若要加密明文 Cipher，首先转换字母 C、i、p、h、e、r 成数字 2、8、15、7、4、17，然后进行加密：

$$5\times\begin{pmatrix}2\\8\\15\\7\\4\\17\end{pmatrix}+\begin{pmatrix}3\\3\\3\\3\\3\\3\end{pmatrix}=\begin{pmatrix}13\\43\\78\\38\\23\\88\end{pmatrix}\bmod 26=\begin{pmatrix}13\\17\\0\\12\\23\\10\end{pmatrix}=\begin{pmatrix}N\\R\\A\\M\\X\\K\end{pmatrix}$$

即在该密钥下，Cipher 经仿射加密后得到的密文是 NRAMXK。

解密：

$$21\times\begin{pmatrix}13\\17\\0\\12\\23\\10\end{pmatrix}-\begin{pmatrix}11\\11\\11\\11\\11\\11\end{pmatrix}=\begin{pmatrix}262\\346\\-11\\241\\472\\199\end{pmatrix}\bmod 26=\begin{pmatrix}2\\8\\15\\7\\4\\17\end{pmatrix}=\begin{pmatrix}C\\I\\P\\H\\E\\R\end{pmatrix}$$

由此可见，原始消息 Cipher 已得到恢复。

2. 多表代替密码

单表代替密码通常其密钥空间很小，因而无法抵抗穷举搜索攻击。此外，它没有将明文字母出现的统计概率掩盖起来，容易遭受频率分析攻击。这里所说的频率分析攻击是指在某种语言中，由于不同字符出现频率的差异所呈现出来的统计规律。

隐藏字母出现的频率分布以及提高代替密码强度的一种方法是采用多个密文字母表，使密文中的每一个字母有多种可能的字母来代替。多表代替密码有多个单字母密钥，每一个密钥被用来加密一个明文字母。第一个密钥加密明文的第一个字母，第二个密钥加密明文的第二个字母，……。在所有的密钥用完后，密钥又再循环使用。

已知明文序列为 $p=p_1p_2\cdots$，$f=f_1f_2\cdots$为映射序列，则对应的密文为：

$$C=E(k,p)=f_1(p_1)f_2(p_2)\cdots \tag{3-7}$$

若 f 是非周期的无限序列，则相应的密码称为非周期多表代替密码。这类密码，对每个明文字母都采用不同的代替表(或密钥)进行加密，称作一次一密密码(One-time Pad Cipher)，这是一种理论上唯一不可破的密码。这种密码对于明文的特点可实现完全隐蔽，但由于需要的密钥量和明文信息的长度相同而限制了它的广泛使用。

在多表代替下，原来明文中的统计特性通过多个表的平均作用而被隐蔽了起来。多表代替密码的破译要比单表代替密码的破译难得多。但是多表代替中的平均结果会使密文的统计特性与明文的统计特性明显不同，随着多表代替周期的加大，这种差别也就更加明显，由此入手就可以破译多表代替密码。

Vigenère 密码、Playfair 密码、滚动密钥密码、弗纳姆密码以及 Hill 密码都属于这一类密码。

1) Vigenère 密码

Vigenère 密码是最著名的多表代替密码，是由法国密码学家 Blaise de Vigenere 于1568 年提出的一种密码，它是一种以移位代替为基础的周期代替密码、多表简单加法密码。Vigenère 密码使用一个词组作为密钥，每一个密钥字母都对应一个代替表。第一个密钥字母用来加密第一个明文字母，第二个密钥字母用来加密第二个明文字母，……，等所有密钥字母都使用完后，密钥又再循环使用。

已知明文 $p=p_1p_2\cdots p_n$，m 为一个固定的正整数，对于一个密钥 $k=k_1k_2\cdots k_m$，则加密算法如下：

$$C=E(p,k)=(p_1+k_1(\mathrm{mod}\ 26),p_2+k_2(\mathrm{mod}\ 26),\cdots,p_i+k_i(\mathrm{mod}\ 26),\cdots) \tag{3-8}$$

解密算法如下：

$$P=D(c,k)=(c_1-k_1(\mathrm{mod}\ 26),c_2-k_2(\mathrm{mod}\ 26),\cdots,c_i-k_i(\mathrm{mod}\ 26),\cdots) \tag{3-9}$$

Vigenère 密码使用 26 个密文字母表，像加法密码一样，他们是一次将明文字母表循环右移 0、1、2、…、25 位的结果。选一个词组或者短语作为密钥，以密钥字母控制使用哪一个密文字母表。

例 3-4　已知明文 p=polyalphabetic cipher，密钥 k=RADIO，即周期 $d=5$，则

明文：p= polyalphabetic cipher

密钥：k= RADIORADIORADI ORADIO。

密文：c= GOOGOCPKIPVTLK QZPKMF。

其中：同一明文字母 p 在不同的位置被加密成不同的字母 G 和 P。

2）Playfair 密码

Playfair 密码将明文中的双字母组合作为一个单元进行处理，并将每一个单元转换成双字母的密文组合。Plairfair 密码基于一个 5×5 矩阵，该矩阵采用一个关键词作为密钥来构造。构造的方法为：按从左至右、从上至下的顺序依次首先填入关键词中非重复的字母，然而再将字母表中剩余的字母按顺序填入矩阵(其中字母 I 和 J 被看作是一个字母)。

对于每一对明文 p_1 和 p_2，其加密方法如下：

(1) p_1 和 p_2 在同一行时，则密文 c_1 和 c_2 分别是紧靠 p_1、p_2 右端的字母。其中第一列看作是最后一列的右方。

(2) 若 p_1 和 p_2 在同一列时，则密文 c_1 和 c_2 分别是紧靠 p_1、p_2 下方的字母。其中第一行看作是最后一行的上方。

(3) 若 p_1 和 p_2 不在同一行也不在同一列时，则密文 c_1 和 c_2 是由 p_1 和 p_2 确定的矩形的其他两角的字母，并且 c_1 和 p_1、c_2 和 p_2 同行。

(4) 若 $p_1=p_2$，则插入一个字符(例如 Q)于重复字母之间。

(5) 若明文字母为奇数时，将空字母 Q 加在明文的末端。

例 3-5　密钥是 EXAMPLE FOR PLAYFAIR，则构造的字母矩阵如表 3-3 所示。

表 3-3　字母矩阵表

E	X	A	M	P
L	F	O	R	Y
I/J	B	C	D	G
H	K	N	Q	S
T	U	V	W	Z

如果明文是 p= chinese student

先将明文每两个分为一组：ch　in　es　es　tu　de　nt

按照加密规则，对应的密文为：IN　CH　PH　PH　UV　IM　HV

Playfair 密码相对于单表代替密码有很大进步，主要体现在两个方面：第一，由于是双字母组合，共有(26×26=676)种组合的可能，识别双字母组合要更为困难；第二，各个字母组合的频率比单字母呈现出大得多的范围，导致频率分析的难度加大。即便如此，Playfair 密码还是相对容易攻破的，因为在密文中仍然存在许多明文语言的结构可被密码分析者利用。

3）滚动密钥密码

对于周期多表代替密码，保密性将随周期 d 的加大而增加，当 d 的长度和明文一样

长时就变成了滚动密钥密码。如果其中所采用的密钥不重复就是一次一密体制。一般地,密钥可取一本书或一篇报告作为密钥源,可由书名、章节号及标题来限定密钥的起始位置。

4）弗纳姆密码

当字母表字母数 $q=2$ 时,滚动密钥密码就变成了弗纳姆密码。

选择随机二元数字序列作为密钥,以 $k=k_1k_2\cdots k_i(k_i\in F_2)$ 表示,其中 F_2 表示只由两个元素构成的二元空间,明文字母编成二元向量后也可以表示为二元序列 $m=m_1m_2\cdots m_i\cdots(m_i\in F_2)$,则加密过程就是将 k 和 m 的相应位逐位地模 2 加,即:

$$c_i = m_i \oplus k_i \quad i = 1,2,\cdots \tag{3-10}$$

译码时,用同样的密钥对密文逐位地模 2 加,便可恢复明文的二元数字序列,即:

$$m_i = c_i \oplus k_i \quad i = 1,2,\cdots \tag{3-11}$$

这种加密方式若使用电子器件实现就是一种序列密码。

5）Hill 密码

Hill 加密算法的基本思想是将 m 个明文字母通过线性变换将它们转换为 m 个密文字母。解密只要做一次逆变换就可以了。密钥就是变换矩阵本身。假设 $m=3$,则

$$\begin{cases} c_1 = k_{11}p_1 + k_{12}p_2 + k_{13}p_3 \\ c_2 = k_{21}p_1 + k_{22}p_2 + k_{23}p_3 \\ c_3 = k_{31}p_1 + k_{32}p_2 + k_{33}p_3 \end{cases} \tag{3-12}$$

可用列向量和矩阵来表示:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{3-13}$$

即加密过程为:

$$C = KP \bmod 26 \tag{3-14}$$

其中,C 和 P 代表密文和明文向量,K 是密钥矩阵。

解密则为:

$$P = K^{-1}C \tag{3-15}$$

例 3-6 加密明文为 july,密钥矩阵为 $k=\begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}$,则加密过程为:

先将明文分为两个组 ju(9,20)和 ly(11,24),加密算法如下:

$$c_1 = \begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} 9 \\ 20 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} 11 \\ 24 \end{bmatrix} = \begin{bmatrix} 11 \\ 22 \end{bmatrix}$$

因此,加密后的密文为 DELW。

解密算法如下:

密钥矩阵的逆矩阵 $k^{-1}=\begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}$,则

$$p_1 = \begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 9 \\ 20 \end{bmatrix}, \quad p_2 = \begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}\begin{bmatrix} 11 \\ 22 \end{bmatrix} = \begin{bmatrix} 11 \\ 24 \end{bmatrix}$$

因此,解密后得到原始密文 july。

6）一次一密密码

一次一密密码是一种理想的加密方案，是由 Major Joseph Mauborgne 和 AT&T 公司的 Gilbert Vernam 在 1917 年发明的。一次一密密码本是一个大的不重复的真随机密钥字母集，这个密钥字母集被写在几张纸上，并一起粘成一个密码本。发送方用密码本中的每一密钥字母准确地加密一个明文字符。加密是明文字符和一次一密密码本密钥字符的模 26 加法。

每个密钥仅对一个消息使用一次。发送方对所发的消息加密，然后销毁密码本中用过的一页或用过的磁带部分。接收方有一个同样的密码本，并依次使用密码本上的每个密钥去解密密文的每个字符。接收方在解密消息后销毁密码本中用过的一页或用过的磁带部分。新的消息则用密码本的新的密钥加密。

密钥字母必须是随机产生的。对这种方案的攻击将是针对用来产生密钥序列的那种方法。使用伪随机序列发生器是不值得考虑的，它们通常具有非随机性。如果采用真随机源，它就是安全的。

另一个重要的事情是密钥序列不能重复使用。一次一密密码本的想法很容易推广到二进制数据的加密，只需由二进制数字组成的一次一密密码本代替由字母组成的一次一密密码，用异或代替一次一密密码本的明文字符加法就成。为了解密，用同样的一次一密密码本对密文异或，其他保持不变，保密性也很完善。

一次一密密码也存在一些缺陷。一方面因为密钥比特必须是随机的，并且绝不能重复使用，密钥序列的长度要等于消息的长度。即使解决了密钥的分配和存储问题，还需确信发送方和接收方是完全同步的。如果接收方有一比特的偏移（或者一些比特在传输过程中丢失了），消息就变成乱的了。另一方面，如果某些比特在传输中被改变了（没有增减任何比特，更像由于随机噪声引起的），那些改变了的比特就不能正确地解密。再者，一次一密密码本不提供鉴别。

3.2.2　置换密码

置换加密算法只把明文中的字母重新排列，字母本身不变，但其位置改变了，这样编成的密码称为置换密码（Permutation Cipher）。最简单的置换密码是把明文中的字母顺序倒过来，然后截成固定长度的字母组作为密文。

例 3-7　明晨 5 点发动反攻

明文：MING　CHENG　WU　DIAN　FA　DONG　FAN　GONG

密文：GNOGN　AFGNO　DAFNA　IDUWG　NEHCG　NIM

这种技巧对密码分析者来说实在微不足道。一种更复杂的方案是把消息一行一行地写成矩形块，然后按列读出，但是把列的次序打乱，列的次序就是算法密钥。

例 3-8　密钥 3 4 2 1 5 6 7。

明文：如表 3-4 所示。

密文：TTNAAPTMTSUOAODWCOIXKNLYPETZ。

单纯的置换密码因为有着与原文相同的字母频率而被识破，如同列变换所示，密码分析可以直接将密文排列成矩阵入手，再来处理列的位置。双字母音节和三字母音节可以派上用场。

表 3-4 例 3-8 的明文字母表

a	t	t	a	c	k	p
o	s	t	p	o	n	e
d	u	n	t	i	l	t
w	o	a	m	x	y	z

多步置换密码相对来说安全得多,但这种复杂的置换是不容易构造出来的。因此,如果例 3-8 中的那条消息用相同算法再加密一次,则密文:NSCYAUOPTTWLTMDNAOIEPAXTTOKZ。

3.3 对称密钥密码技术

在很长一段时间里,密码技术主要应用于军事以及外交等领域,直到 1977 年美国国家标准局公布实施了“美国数据加密标准(DES)”,军事部门垄断密码的局面才被打破,民间力量开始全面介入密码学的研究和应用中。市场上涌现出大量的民用加密产品,常用的加密算法有 DES、IDEA、AES 等。

对称加密的基本要求:

(1) 需要强大的加密算法。加密算法至少应该满足:即使分析人员知道了算法并能访问一些或者更多的密文,也不能破译出密文或者得出密钥。这个要求若以更强硬的形式表达出来,那就是:即使分析人员拥有一些密文和生成密文的明文,也不能译出密文或者发现密钥,即加密算法应足以抵抗已知明文类型的破译。

(2) 发送方和接收方必须用安全的方式来获得密钥的副本,以保证密钥的安全。如果有人发现了密钥,并知道了算法,则使用此密钥的所有通信便都是可读取的。

对称密钥密码技术有两种不同的实现方式,分别是流密码技术和分组密码技术。下面将对这两种典型的密码技术加以介绍。

3.3.1 流密码技术

流密码的基本思想是利用密钥 k 产生一个密钥流 $k_0k_1k_2\cdots$,并使用如下规则对明文 $p=p_0p_1p_2\cdots$加密:$c=c_0c_1c_2\cdots=E_{k0}(p_0)E_{k1}(p_1)E_{k2}(p_2)\cdots$。密钥流由密钥流发生器 f 产生:$z_i=f(k,\sigma_i)$,这里 σ_i 是加密器中的记忆元件(存储器)在时刻 i 的状态,f 是由密钥 k 和 σ_i 产生的函数。

流密码的滚动密钥 $z_0=f(k,\sigma_0)$由函数 f、密钥 k 和指定的初态 σ_0 完全确定。此后,由于输入加密器的明文可能影响加密器中内部记忆元件的存储状态,因此 $\sigma_i(i>0)$可能依赖于 k、σ_0、x_0、x_1、…、x_{i-1}等参数。

根据加密器中记忆元件的存储状态 σ_i 是否依赖于输入的明文字符,流密码可进一步分成同步和自同步两种。σ_i 独立于明文字符的叫做同步流密码,否则叫做自同步流密码。由于自同步流密码的密钥流的产生与明文有关,因而较难从理论上进行分析。目前大多数研

究成果都是关于同步流密码的。在同步流密码中，由于 $z_i=f(k,\sigma_i)$ 与明文字符无关，因而此时密文字符 $y_i=E_{Z_i}(x_i)$ 也不依赖于此前的明文字符。因此，可将同步流密码的加密器分成密钥流产生器和加密变换器两个部分。如果与上述加密变换对应的解密变换为 $x_i=D_{Z_i}(y_i)$，则可给出同步流密码体制的模型如图 3-2 所示。

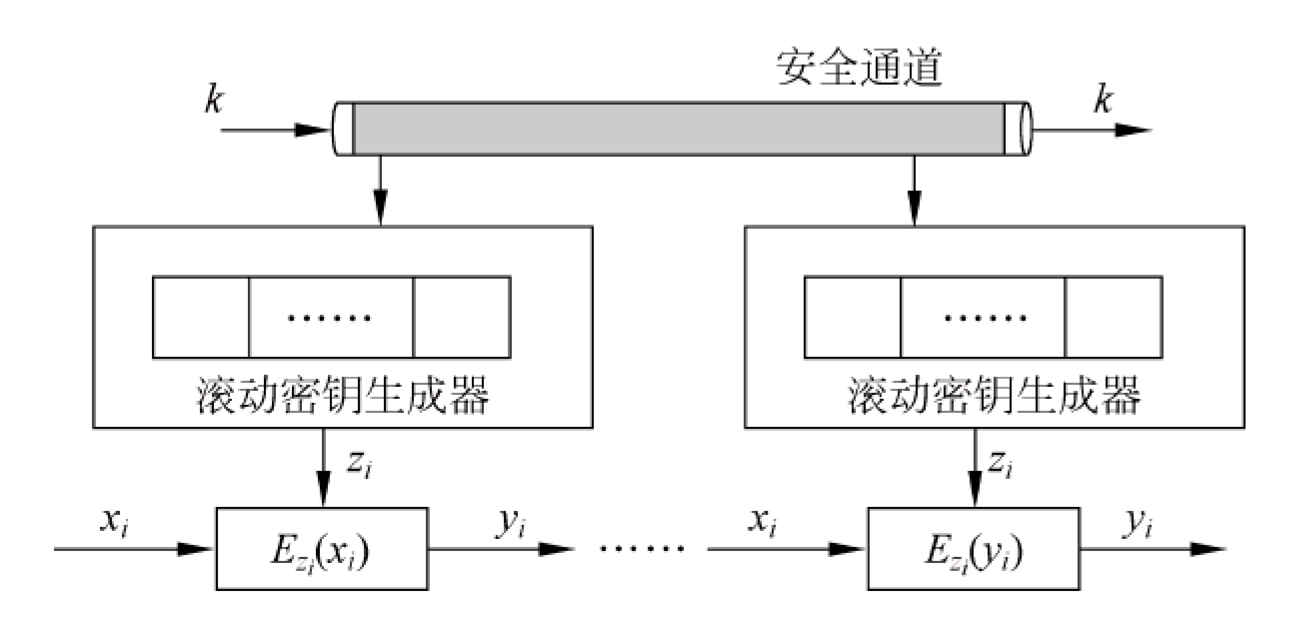


图 3-2　同步流密码体制的模型

实际使用的数字保密通信系统一般都是二元系统，因而在有限域 GF(2) 上讨论的二元加法流密码是目前最为常用的流密码体制，其加密变换可表示为 $y_i=z_i\oplus x_i$。实际使用中，密码设计者的最大愿望是设计出一个滚动密钥生成器，使得密钥经其扩展成的密钥流序列具有如下性质：极大的周期、良好的统计特性、抗线性分析和抗统计分析。

下面将详细介绍两种流密码算法：A5/1 和 RC4。这两种算法在当今被广泛应用。A5/1 在 GSM 移动通信中使用，A5/1 算法是基于硬件实现的流密码的代表。RC4 算法在安全套接字 SSL 协议等许多地方有广泛的使用，是一种特殊的流密码，其软件实现效率非常高。

1. A5/1

A5/1 算法主要应用在 GSM 移动通信中用于保护数据。该算法可以通过代数描述，也可任意使用简单的流程图来描述。在这里同时给出这两种描述。

A5/1 使用 X、Y、Z 3 个线性移位寄存器 LFSR。寄存器 X 包括 19 比特，编号为 $(x_0, x_1\cdots x_{18})$。寄存器 Y 包括 22 比特，编号为 $(y_0,y_1\cdots y_{21})$。寄存器 Z 包括 23 比特，编号为 $(z_0,z_1\cdots z_{22})$。3 个 LFSR 总共包括 64 比特。

密钥 k 同样是 64 比特，用于初始化 3 个寄存器。用密钥填充 3 个寄存器后，就完成了密码流生成前的准备。在描述密码流之前，首先介绍 3 个寄存器的详细结构。

对于寄存器 X，每步进行如下操作：

$$\begin{aligned} t &= x_{13}\oplus x_{16}\oplus x_{17}\oplus x_{18} \\ x_i &= x_{i-1},\quad i=18、17、16、\cdots、1 \\ x_0 &= t \end{aligned} \tag{3-16}$$

类似地，对于寄存器 Y 和 Z，每步分别进行如下操作：

$$\begin{aligned} t &= y_{20}\oplus y_{21} \\ y_i &= y_{i-1},\quad i=21、20、19、\cdots、1 \\ y_0 &= t \end{aligned} \tag{3-17}$$

和

$$t = z_7 \oplus z_{20} \oplus z_{21} \oplus z_{22}$$
$$z_i = z_{i-1}, \quad i = 22、21、20、\cdots、1$$
$$z_0 = t \tag{3-18}$$

给定 3 个比特 x、y、z，定义 maj(a,y,z)为“多数投票”函数，即如果 x、y、z 中的多数为 0，则函数返回 0，否则返回 1。

A5/1 使用硬件实现，每个时钟周期作如下计算：

$$m = \text{maj}(x_8, y_{10}, z_{10}) \tag{3-19}$$

于是寄存器 X、Y、Z 依照如下规则进行处理：

(1) 如果 $x_8 = m$，那么就进行 X 操作。

(2) 如果 $y_{10} = m$，那么就进行 Y 操作。

(3) 如果 $z_{10} = m$，那么就进行 Z 操作。

最后，密钥流比特 s 按照如下关系产生：

$$s = x_{18} \oplus y_{21} \oplus z_{22} \tag{3-20}$$

为了生成一个比特的密钥流的过程看似复杂，但由于 A5/1 的硬件实现非常简单，比特产生的速度与时钟速度相当。并且从一个 64 位的密钥可以产生无穷多的密钥流，尽管最终密钥流将出现循环。A5/1 算法可以使用简单的电码表示，如图 3-3 所示。

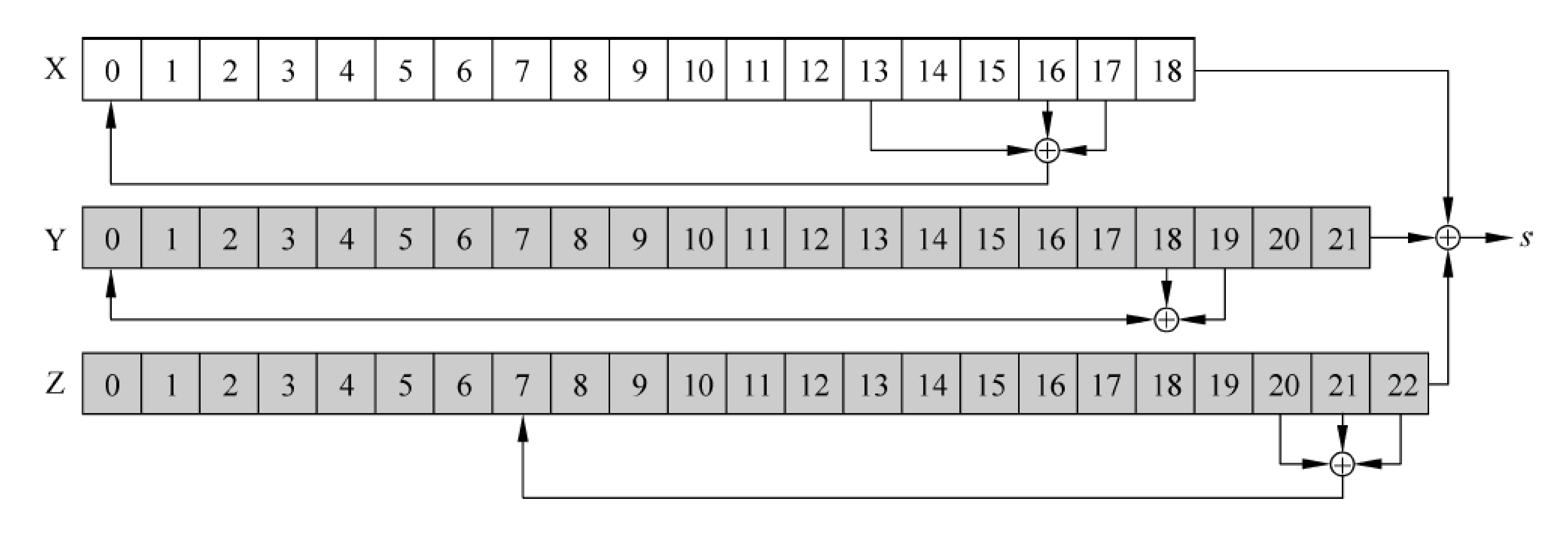


图 3-3 A5/1 密钥流生成

2. RC4

RC4 也是一种流密码，但是它与 A5/1 有很大的不同。RC4 算法专门为软件实现优化，而 A5/1 则是根据硬件实现设计；RC4 每步产生一个密钥字节，而 A5/1 每步仅产生一个密钥流比特。

RC4 算法非常简单，因为从本质上来讲它就是一个包含 256 字节的置换查找表，在产生密钥流的每一个字节时，所查的表就进行一次修改，表始终都包含{0,1,2,…,255}的置换。

整个 RC4 算法都是基于字节的。算法的第一阶段是对于查表使用的密钥进行初始化，用 key[i]表示密钥，这里 i=1、2、…、$N-1$，每个 key[i]是一个字节，标记为 s[i]，这里每个 s[i]也是一个字节。置换 S 的初始化过程的代码如表 3-5 所示。RC4 的一个特点是，密钥可以是 0 到 256 字节的任意长度。密钥只在初始化置换 S 中使用。

初始化阶段完成后，通过表 3-6 中的算法产生每个密钥流字节。用 keystreamByte 表

示输出，在加密时与明文做 XOR 运算，解密时与密文做 XOR 运算。RC4 算法的输出同时也可作为伪随机序列生成器使用。

表 3-5　RC4 初始化

```
for i = 0 to 255
    s[i] = i
    k[i] = key[i mod N]
Next i
j = 0
for i = 0 to 255
    j = (j + s[i] + k[i]) mod 256
    swap(s[i], s[j])
next i
i = j = 0
```

表 3-6　RC4 密钥流字节

```
i = (i + 1) mod 256
j = (j + s[i]) mod 256
swap(s[i], s[j])
t = (s[i] + s[j]) mod 256
keystreamByte = s[t]
```

RC4 算法可以被视为自修改的查找表，它非常简单，并且软件实现效率很高。然而，对于 RC4 存在可行的攻击方法，但是只要在使用时丢弃生成前 256 字节密钥流，该攻击就不可行。这可以通过在初始化过程中额外添加 256 步来完成，每一步产生表 3-6 中被丢弃的密钥流字节。

3.3.2　分组密码技术

分组密码是对称密码的典型代表。即数据在密钥的作用下，一组一组地被处理，并且通常明文和密文的长度是相等的。一次对一个明文分组（例如 DES 为 64 位）进行加密，而且每次的加密密钥都相同。分组加密的一般结构如图 3-4 所示。

当密钥给定时，对于每一个明文分组，都有唯一的一个密文分组与之对应。因此可以想象有一个非常大的电码本，对每一个可能的明文分组，在电码本中都有唯一与之对应的密文分组。对于大于分组长度的报文，需将其分为若干特定分组长度的分组，最后一个分组可能需要填充。解密过程也是一次对一个密文分组进行解密。而且每次解密都使用同一个密钥。

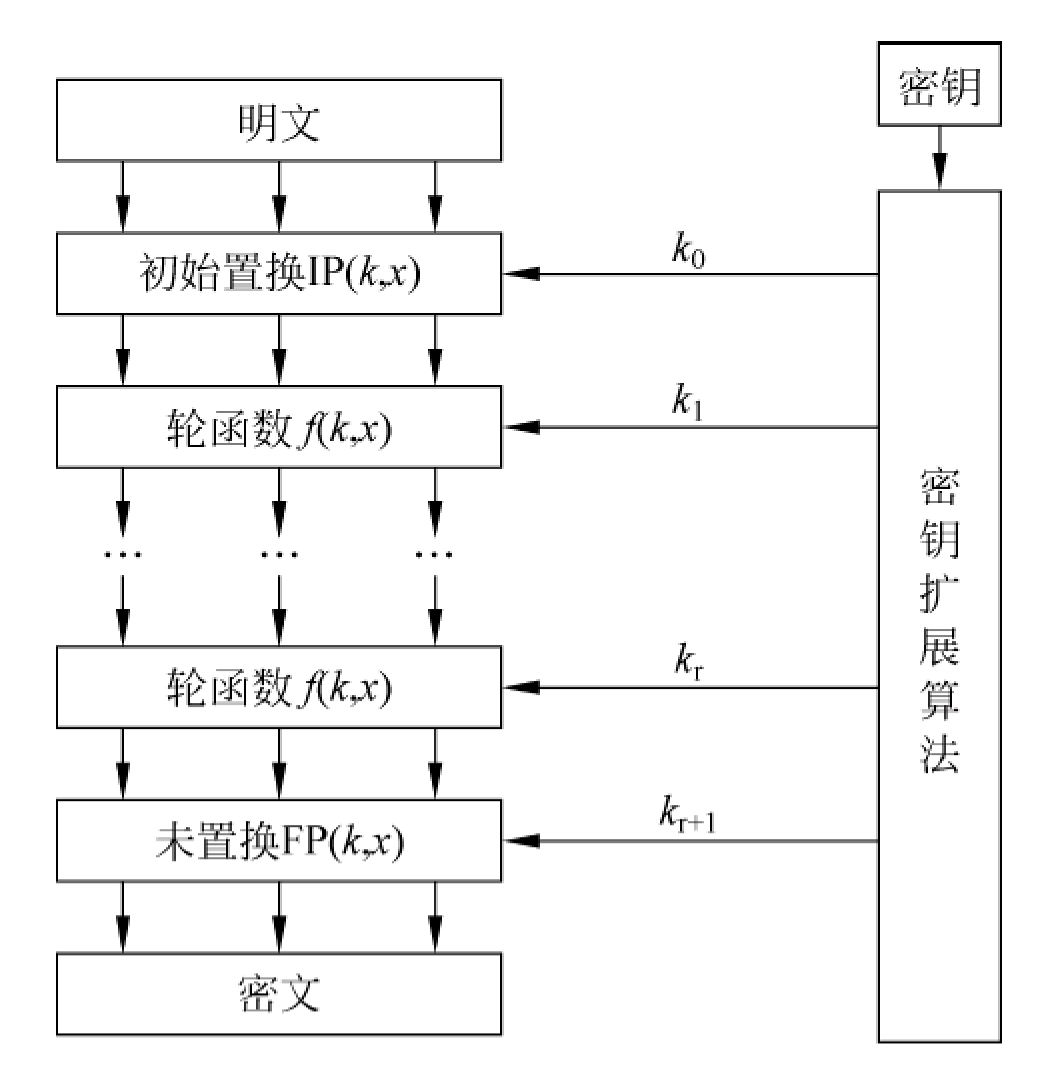


图 3-4　分组加密的一般结构

用于短数据（例如加密密钥）加密时效果非常理想。但如果同一明文分组在消息中反复出现，产生的密文分组就会相同，不仅容易被攻击者抓住规律猜测攻击，而且在时间上也大大重复了相同的工作。因此，用于长消息时安全性不够。

给定加密消息的长度是随机的，按特定长度（例如 64 bit）分组时，最后一组消息长度可能不足 64bit，如图 3-5 所示。这时可以填充一些数字，通常用最后一个字节作为填充指示符（PI），它所表示的十进制数字就是填充占有的字节数。数据尾部、填充字符和填充指示符

一起作为一组进行加密。

1．数据加密标准（DES）

DES(Data Encryption Standard）是一种分组乘积密码，包括 16 轮迭代。明/密文分组长度为 64 位，密钥总长为 64 位，有效长度为 56 位，其中第 8、16、…、64 位共 8 位是奇偶校验位。DES 是一种对称运算，除子密钥使用顺序逆序外，加密和解密算法相同。DES 是一种面向二进制的密码算法，能够加/解密任何形式的计算机数据。

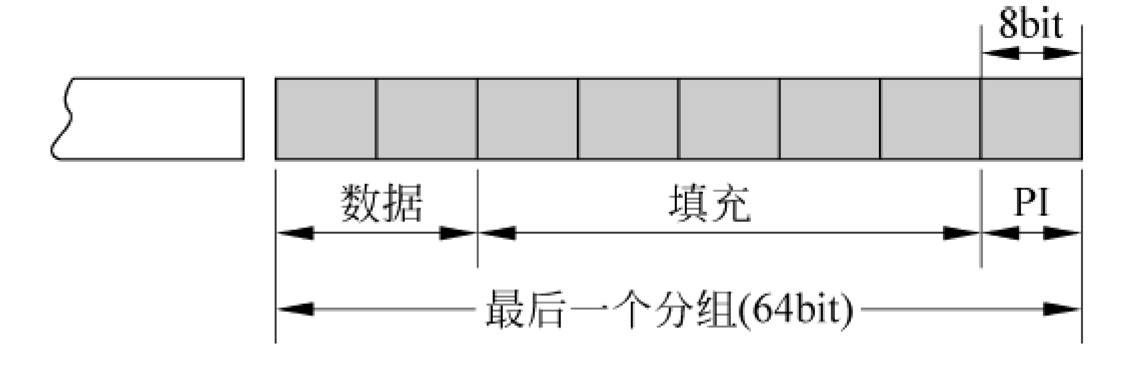


图 3-5　分组密码技术

DES 的加密算法流程如图 3-6 所示，主要包括三大步骤：

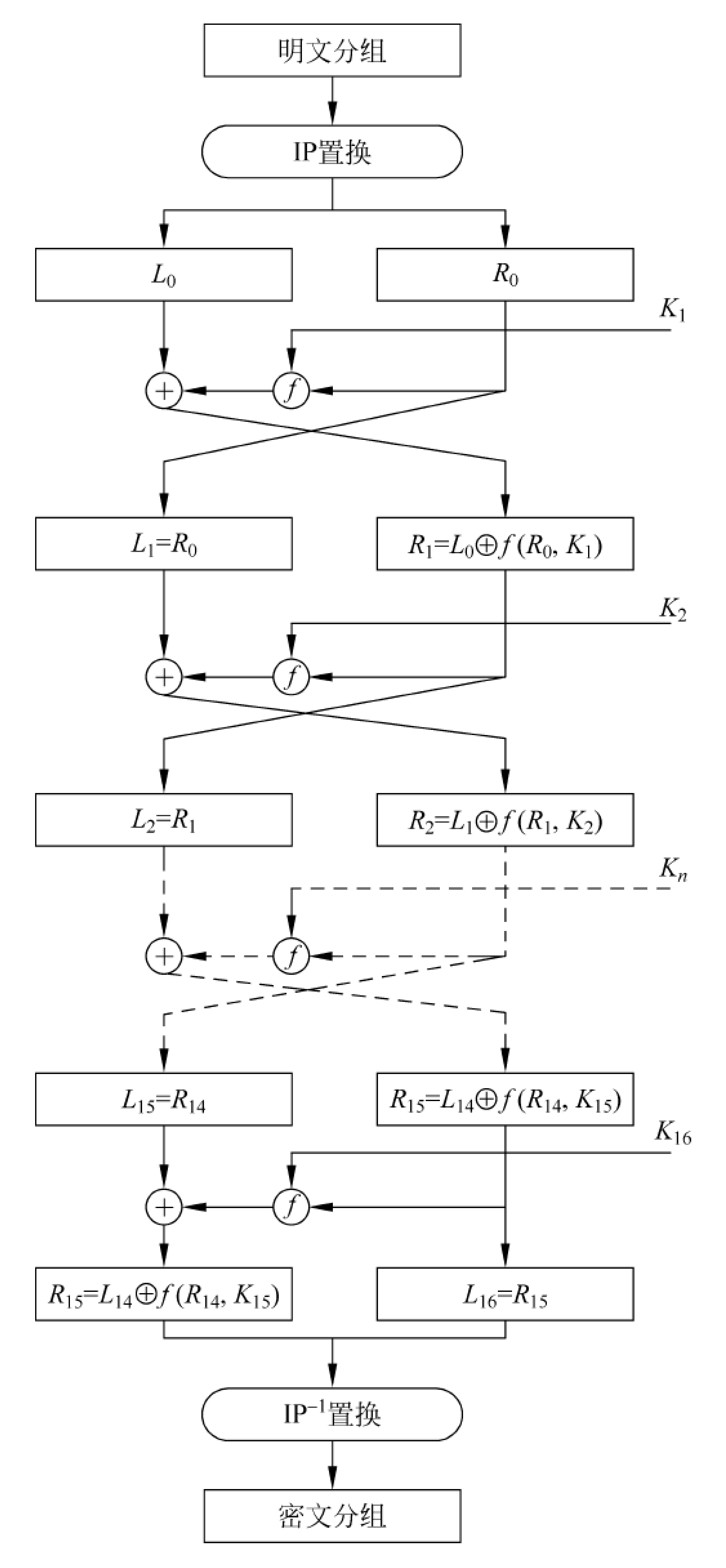


图 3-6　DES 加密算法流程图

(1) 初始置换 IP：把输入的 64 位数据块的排列顺序打乱，每位数据按照下面的换位规则重新组合：IP $(b_1b_2b_3\cdots b_{64})=b_{58}b_{50}\cdots b_7$，即将输入的第 58 位换到输出的第 1 位，将输入的第 50 位换到输出的第 2 位，……，输入的第 7 位换到输出的第 64 位，将变换后的数据平分成各 32 位的左右两部分，左部分记为 L_0，右部分记为 R_0，如表 3-7 所示。

表 3-7　IP 置换

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

(2) 16 轮的轮变换：首先密钥扩展算法将 64 位的输入密钥(称为主密钥 Master Key)扩展为加/解密各轮所需的轮子密钥(Sub Key)。DES 共需要 16 个轮子密钥，每个轮子密钥有 48 位。对 R_0 实行在轮子密钥 k_1(轮子密钥由密钥扩展算法产生)控制下的变换 f，结果记为 $f(R_0,k_1)$，再与 L_0 做按位异或运算，其结果记为 R_1，R_0 则直接作为下一轮的 L_1，如此循环 16 轮，得到预输出结果 R_{16}、L_{16}。

$$\begin{cases} L_n = R_{n-1} \\ R_n = L_{n-1} \oplus f(R_{n-1},K_n) \end{cases},\quad n=1,2,\cdots,16 \tag{3-21}$$

f 函数是多个置换函数和替代函数的组合函数，它将 32 位比特的输入变换为 32 位的输出。如图 3-6 所示，32 位的 R 经过扩展变换 E 后，扩展为 48 位的 E(R)，然后与 48 位的轮子密钥 k 进行按位异或。$\mathrm{E}(R)=\mathrm{E}(b_1b_2b_3\cdots b_{32})=b_{32}b_1\cdots b_1$，输出的第 1 位为输入的第 32 位，输出的第 2 位为输入的第 1 位，输出的第 48 位为输入的第 1 位，如表 3-8 所示。E 的主要作用是增加算法的扩散效果。

表 3-8　E 置换

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

(3) 逆初始置换 IP^{-1}：逆初始置换 IP^{-1}是初始置换 IP 的逆置换，它将由 L_{16}、R_{16}合并的 64 位数据作为输入，进行换位后得到 64 位的密文输出。$\mathrm{IP}^{-1}(b_1b_2b_3\cdots b_{64})=b_{40}b_8\cdots b_{25}$，即将输入的第 40 位换到输出的第 1 位，将输入的第 8 位换到输出的第 2 位，……，输入的第 25 位换到输出的第 64 位，如表 3-9 所示。

表 3-9 IP^{-1}置换

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
37	6	46	14	54	22	62	30
36	5	45	13	53	21	61	29
35	4	44	12	52	20	60	28
34	3	43	11	51	19	59	27
33	2	42	10	50	18	58	26
32	1	41	9	49	17	57	25

2. TDEA 和 IDEA

1) TDEA 算法

TDEA(Triple Data Encryption Algorithm)算法又叫做三重 DES 算法，它需要执行 3 次 DES 的加密，如图 3-7 所示。一般三重 DES 算法使用两个 DES 密钥。TDEA 算法的加密步骤如下：

(1) 发送端用密钥 Key_1进行 DES 加密。

(2) 发送端用密钥 Key_2 对上一结果进行 DES 解密。

(3) 发送端用密钥 Key_3 对上一结果进行 DES 加密。

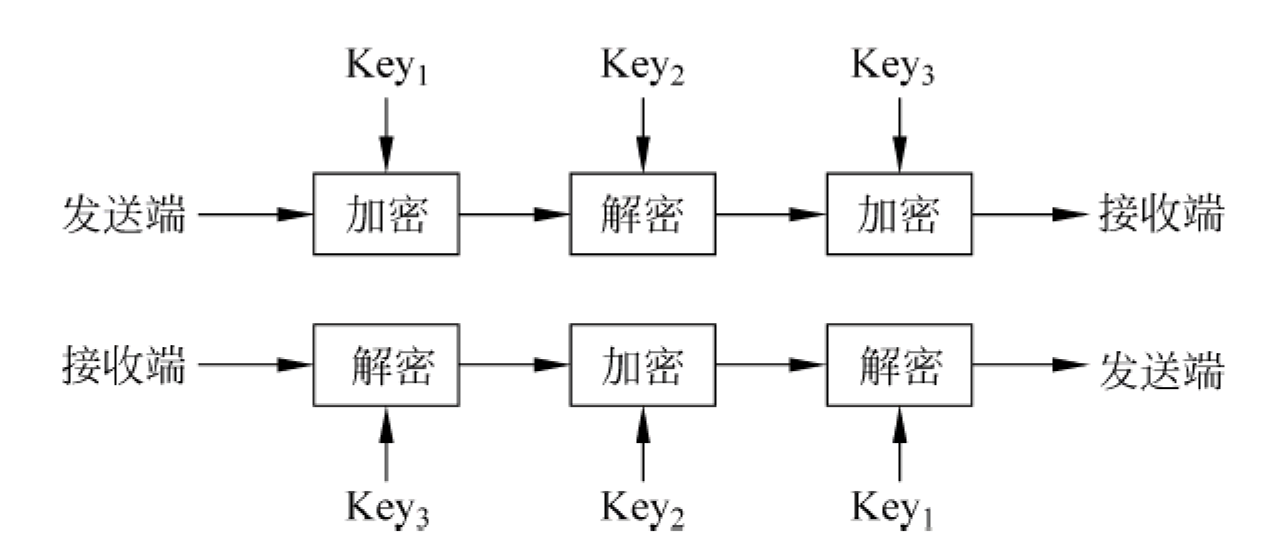


图 3-7 TDEA 算法的加/解密过程

2) IDEA 算法

IDEA(Internatinal Data Encryption Algorithm，国际数据加密算法)算法是一种对称分组密码算法，是由瑞士联邦理工学院 Xuejia Lai 和 James Massey 在 1990 年提出的。IDEA 加密算法是在 DES 算法的基础上发展而来的，类似于三重 DES 算法，其分组长度也是 64 位，但密钥长度是 128 位。IDEA 算法是用 128 位密钥对 64 位二进制码组成的数据组进行加密的，也可用同样的密钥对 64 位密文进行解密变换。IDEA 的密钥比 DES 的多一倍，增加了破译难度。

IDEA 算法也是通过一系列的加密轮次操作的，每轮加密都使用从完整的加密密钥生成的一个子密钥。IDEA 使用的运算有异或、模 2^{16}加法和模($2^{16}+1$)乘法。这 3 种运算彼此混合可产生很好的效果。IDEA 整个算法包含子密钥产生、数据加密、数据解密 3 个部分，其基本工作原理如图 3-8 所示。

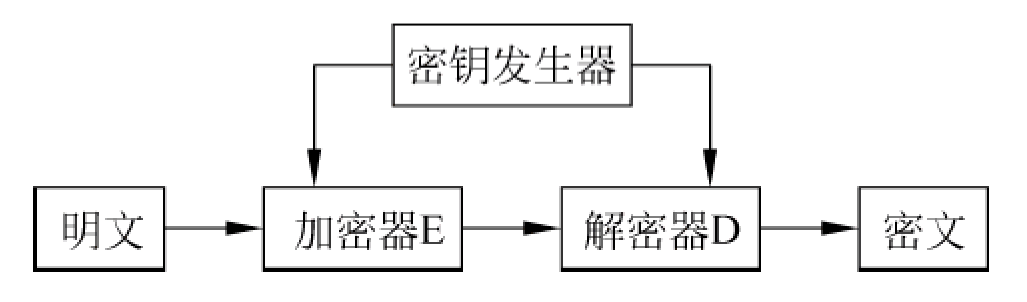


图 3-8 IDEA 算法的工作原理

IDEA 算法的加/解密过程可归纳如下：

(1) 8 轮的重复运算。

(2) 64 位的明文分组在每一轮都是被分成 4 个子分组,每个 16 位子分组作为一个单元来处理。

(3) 每一轮中有 6 个不同的子密钥参与运算。

(4) 最后的输出变换有 4 个子密钥参与运算。

IDEA 的解密算法使用与加密算法同样的结构,是一个将密文分组当作输入而逐步恢复明文分组的解密过程。与加密过程不同的是子密钥的生成方法。

(1) 解密循环 i 的前 4 个子密钥从加密循环$(10-i)$的前 4 个子密钥中导出;解密密钥第 1、4 个子密钥对应于 1、4 加密子密钥的乘法逆元;解密密钥第 2、3 个子密钥对应于 2、3 加密子密钥的加法逆元。

(2) 对前 8 个循环来说,循环 i 的最后两个子密钥等于加密循环$(9-i)$的最后两个子密钥。

3. 高级加密标准(AES)

DES 的 56 比特密钥实在太小,虽然三重 DES 可以解决密钥长度的问题,但是 DES 的设计主要针对硬件实现,而在当今许多领域,需要用软件方法来实现它,在这种情况下,它的效率相对较低。鉴于此,1997 年 4 月 15 日美国国家标准和技术研究所(NIST)发起征集 AES(Advanced Encryption Standard,高级加密标准)算法的活动,并成立了 AES 工作组,目的是为了确定一个非保密的、公开披露的、全球免费使用的加密算法,用于保护下一世纪政府的敏感信息,也希望能够成为保密和非保密部门公用的数据加密标准。

AES 是 Rijndael 算法的一个子集,其算法是 128 位块密码,支持 3 种不同大小的密钥:128 位、192 位和 256 位。最大优点是可以给出算法的最佳差分特征的概率及最佳线性逼近的偏差的界,由此可以分析算法抵抗差分密码分析及线性密码分析的能力。

AES 密码算法采用的是代替-置换网络(Substitution-Permutation Network,SPN)结构,每一轮操作由 4 层组成:第 1 层(字节替换)为非线性层,用 S 盒(见后文中的介绍)对每一轮中的单个字节分别进行替换;第 2 层(行移位)和第 3 层(列混合)是线性混合层,对当前的状态按行移位,按列混合;第 4 层(密钥加层)用子密钥与当前状态进行字节上的异或。AES 的具体算法结构如图 3-9 所示。

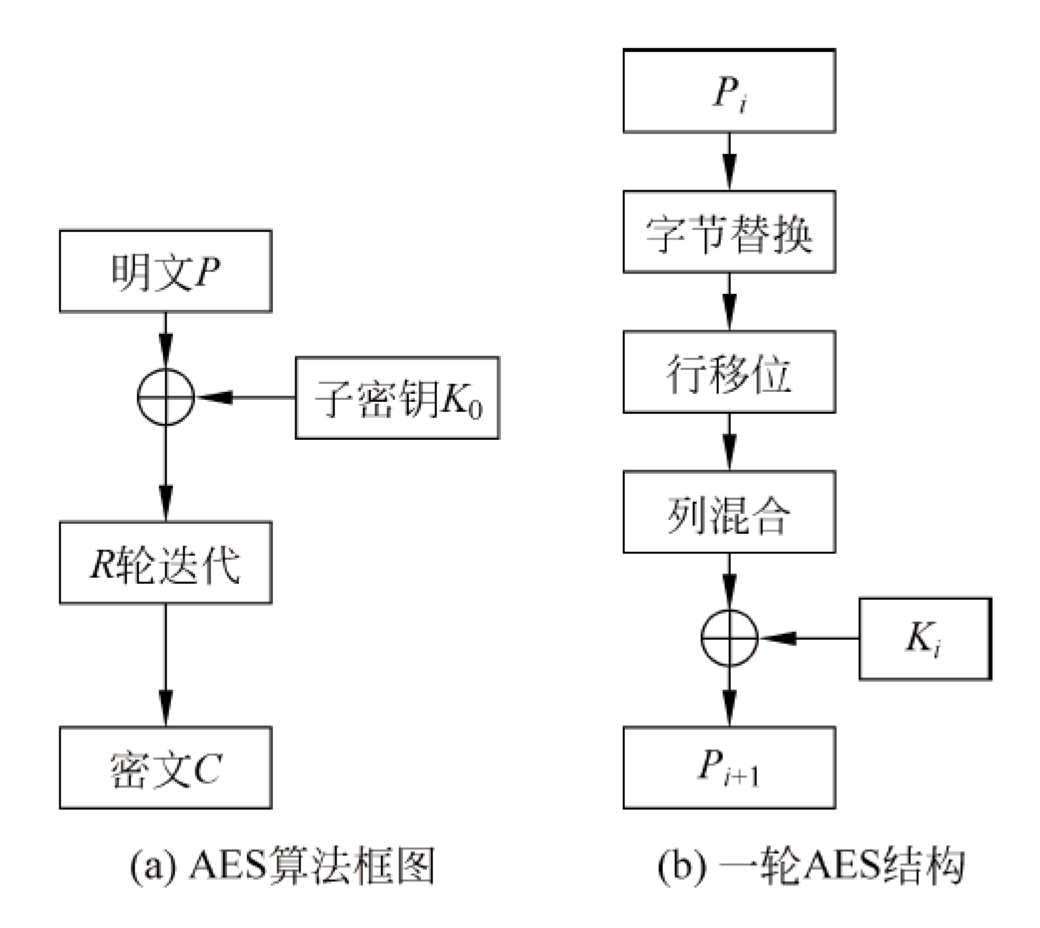


图 3-9　AES 算法结构图

图 3-9 中，(a)图给出了算法的整体结构，输入明文 P 与子密钥 K_0 异或，然后经过 R 轮迭代最终生成密文 C，其中第 1 到($R-1$)轮迭代结构为图(b)所示，第 R 轮与前面各轮稍微有点不同，缺少混合层。

其中，加密轮数与密钥长度的关系如表 3-10 所示。

表 3-10 AES 参数

密钥长度(bit)	128	192	256
明文分组长度(bit)	128	128	128
轮数	10	12	14
每轮密钥长度(bit)	128	128	128
扩展密钥长度(B)	176	206	240

1) 字节替换(SubBytes)

AES 定义了一个 S 盒(即 State)，State 中的每个字节按照如下方式映射为一个新的字节：把该字节的高 4 位作为行值，低 4 位作为列值，然后取出 S 盒中对应行和列的元素作为输出。例如，十六进制数{84}，对应 S 盒的行是 8 列是 4，S 盒中该位置对应的值是{5F}。

S 盒是一个由 16×16 字节组成的矩阵，包含了 8 位值所能表达的 256 种可能的变换。S 盒按照以下方式构造：

(1) 逐行按照升序排列的字节值初始化 S 盒。第一行是{00}、{01}、{02}、…、{OF}；第二行是{10}、{l1}、…、{1F}等。在行 x 和列 y 的字节值是{xy}。

(2) 把 S 盒中的每个字节映射为它在有限域 GF(2^k)中的逆。GF 代表伽罗华域，GF(2^8)由一组从 0x00 到 0xff 的 256 个值组成，加上加法和乘法。

$$GF(2^8)=\frac{Z_2[X]}{(x^8+x^4+x^3+x+1)}$$

其中 $Z_2[X]$是字节值的二进制表示形式。{00}被映射为它自身{00}。

(3) 把 S 盒中的每个字节记成($b_7,b_6,b_5,b_4,b_3,b_2,b_1,b_0$)。对 S 盒中每个字节的每位做如下变换：

$$b'_i=b_i\oplus b_{(i+4)\bmod 8}\oplus b_{(i+5)\bmod 8}\oplus b_{(i+6)\bmod 8}\oplus b_{(i+7)\bmod 8}\oplus c_i \tag{3-22}$$

式(3-22)中，c_i 是指值为{63}字节 c 的第 i 位，即($c_8,c_7,c_6,c_5,c_4,c_3,c_2,c_1,c_0$)=(01100011)。符号(′)表示更新后的变量的值。AES 用以下的矩阵方式描述了这个变换：

$$\begin{bmatrix}b_0\\b_1\\b_2\\b_3\\b_4\\b_5\\b_6\\b_7\end{bmatrix}=\begin{bmatrix}1&0&0&0&1&1&1&1\\1&1&0&0&0&1&1&1\\1&1&1&0&0&0&1&1\\1&1&1&1&0&0&0&1\\1&1&1&1&1&0&0&0\\0&1&1&1&1&1&0&0\\0&0&1&1&1&1&1&0\\0&0&0&1&1&1&1&1\end{bmatrix}\begin{bmatrix}b_0\\b_1\\b_2\\b_3\\b_4\\b_5\\b_6\\b_7\end{bmatrix}+\begin{bmatrix}1\\1\\0\\0\\0\\1\\1\\0\end{bmatrix}$$

2) 行位移变换(ShiftRows)

State 的第 1 行字节保持不变，State 的第 2 行字节循环左移一个字节，State 的第 3 行

字节循环左移两个字节，State 的第 4 行循环左移 3 个字节，如图 3-10 所示。

14	0	5d	ab
78	10	C1	fd
31	9	11	3f
28	0b	2a	45

ShiftRows 变换 →

14	0	5d	ab
10	C1	fd	78
11	3f	31	9
45	28	0b	2a

图 3-10　ShiftRows 变换

3）列混合变换（MixColumns）

列混合变换是一个替代操作，是 AES 最具技巧性的部分。它只在 AES 的第 0、1、…、$(R-1)$轮中使用，在第 R 轮中不使用该变换。乘积矩阵中的每个元素都是一行和一列对应元素的乘积之和。在 MixColumns 变换中，乘法和加法都是定义在 $GF(2^8)$上的。State 的每一列 $(b_{i,j})$，$i=0、1、2、3$，$j=0、1、\cdots、L_{b-1}$，L_b 为分组长度 4words（128bit），被理解为 $GF(2^8)$上的多项式，该多项式与常数多项式 $a(x)=a_3x^3+a_2x^2+a_1x+a_0$ 相乘并模 $M(x)=x^4+1$ 约化。

这个运算需要做 $GF(2^8)$上的乘法。但由于所乘的因子是 3 个固定的元素 02、03、01（在进行具体运算时每行有固定的排列顺序），所以这些乘法运算仍然是比较简单的（注意到乘法运算所使用的模多项式为 $m(x)=x^8+x^4+x^3+x+1$）。设一个字节为 $b=(b_7b_6b_5b_4b_3b_2b_1b_0)$，则：

$$b\times\{01\}=b$$
$$b\times\{02\}=b_6b_5b_4b_3b_2b_1b_0 0$$
$$b\times\{03\}=b\times\{01\}+b\times\{02\}$$

其中，加法为取模 2 的加法，即逐比特异或。

写成矩阵形式为：

$$\begin{bmatrix}b_0\\b_1\\b_2\\b_3\end{bmatrix}=\begin{bmatrix}02&03&01&01\\01&02&03&01\\01&01&02&03\\03&01&01&02\end{bmatrix}\begin{bmatrix}a_0\\a_1\\a_2\\a_3\end{bmatrix}$$

4）密钥加变换（Add RoundKey）

Add RoundKey 称为轮密钥加变换。128 位的 State 按位与 128 位的密钥 XOR：

$$(b_{0j},b_{1j},b_{2j},b_{3j})\leftarrow(b_{0j},b_{1j},b_{2j},b_{3j})\oplus(k_{0j},k_{1j},k_{2j},k_{3j})$$

对 $j=0,1,\cdots,(R-1)$轮密钥加变换很简单，却影响了 State 中的每一位。密钥扩展的复杂性和 AES 的其他阶段运算的复杂性确保了该算法的安全性。

5）密钥扩展（Key Expansion）

为了防止已有的密码分析攻击，AES 使用了与轮相关的轮常量 Rcon[j]（是一个字，这个字的右边 3 个字节总为 0）防止不同轮中产生的轮密钥的对称性或相似性。AES 在加密和解密算法中使用了一个由种子密钥字节数组生成的密钥调度表，AES 规范中称为密钥扩展（Key Expansion）。密钥扩展过程从一个原始密钥中生成多重密钥以代替使用单个密钥，大大增加了比特位的扩散，在 AES 密钥扩展算法的输入值是 4 字密钥，输出是一个 44 字的

一维线性数组。这足以为初始轮密钥扩展过程阶段和算法中的其他 10 轮中的每一轮提供 16 字节的轮密钥。

通过生成器产生 N_r+1 轮轮密钥，每个轮密钥由 N_b 个字组成，共有 $N_b(N_r+1)$个字 $w[i]$，$i=0、1、\cdots、N_b(N_r+1)-1$。

在加密过程中，需要 N_r+1 个子密钥，需要构造 $4(N_r+1)$个 32 位字。Rijndael 的密钥扩展方案的代码描述如下：

```
KeyExpansion(Byte key[4Nk],word w[Nb(Nr+1)],Nk)
{//Nk 代表以 32 位字为单位的密钥的长度，即 Nk = 密钥长度/32begin
  i = 0
  while(i<Nk)
  w[i] = word[key[4i],key[4i+1],key[4i+2],key[4i+3]]
  i = i+1
end while
i = Nk
while(i<Nb(Nr+1))
   word temp = w[i-1]
   if (i mod Nk = 0)
     temp = SubWord(RotWord(temp)) xor Rcon[i/Nk]
   else if (Nk = 8 and i mod Nk = 4)
     temp = SubWord(temp)
end if
w[i] = w[i-Nk] xor temp
i = i+1
end while
end
}
```

其中，key[]和 w[]分别用于存储扩展前和扩展后的密钥。SubWord()、RotWord()分别是与 S 盒的置换和以字节为单位的循环移位。Rcon[i]=(RC[i],'00','00','00')，RC[0]='01'，RC[i]=2·(RC[$i-1$])。前 10 个轮常数 RC[i]的值(用十六进制表示)如表 3-11 所示，其对应的 Rcon[i]如表 3-12 所示。

表 3-11　RC[*i*]

i	1	2	3	4	5	6	7	8	9	10
RC[*i*]	01	02	04	08	10	20	40	80	1b	36

表 3-12　Rcon[*i*]

i	1	2	3	4	5
Rcon[*i*]	01000000	02000000	04000000	08000000	10000000
i	6	7	8	9	10
Rcon[*i*]	20000000	40000000	80000000	16000000	36000000

输入密钥直接被复制到扩展密钥数组的前 4 个字中，得到 $w[0]、w[1]、w[2]、w[3]$；然后每次用 4 个字填充扩展密钥数组余下的部分。在扩展密钥数组中，$w[i]$的值依赖于

$w[i-1]$和 $w[i-4]$($i \geqslant 4$)。

对 w 数组中下标不为 4 的倍数的元素，只是简单地异或，其逻辑关系为：$w[i]=w[i-1] \oplus w[i-4]$($i$ 不为 4 的倍数)。

对 w 数组中下标为 4 的倍数的元素，采用如下的计算方法：

(1) 将一个字的 4 个字节循环左移一个字节，即将字$[b_0, b_1, b_2, b_3]$变为$[b_1, b_2, b_3, b_0]$。

(2) 基于 S 盒对输入字中的每个字节进行 S 代替。

(3) 将步骤(1)和步骤(2)的结果再与轮常量 Rcon[i]相异或。

3.3.3　对称密钥密码的分析方法

密码编码学和密码分析学既对立又统一，正是由于它们的对立性才促进了密码学的发展。密码分析学是在不知道密钥的情况下，恢复出密文中明文信息的方法。根据密码破译者对明文、密文等信息掌握的多少，可以将密码分析分为以下 5 种情形：

(1) 唯密文攻击：对于该形式的密码分析，破译者只知道加密算法和待破译的密文。

(2) 已知明文攻击：破译者已知的内容包括加密算法和经密钥加密形成的一个或多个明文-密文对。

(3) 选择明文攻击：破译者除了知道加密算法外，他还可以选定明文消息，并可以知道对应加密得到的密文。

(4) 选择密文攻击：破译者除了知道加密算法外，还包括他自己选定的密文和对应的、已解密的明文。

(5) 选择文本攻击：破译者已知的东西包括加密算法、由密码破译者选择的明文消息和它对应的密文，以及由密码破译者选择的猜测性明文和它对应的已破译的明文。

具体的分析方法主要包括：

1. 强力攻击法

强力攻击法可用于任何分组密码，且攻击的复杂度仅依赖于分组长度和密钥长度。严格地讲，攻击所需的时间复杂度依赖于分组密码的工作效率，其工作效率包括加/解密速度、密钥扩展速度和存储空间等。

2. 差分密码分析

差分密码分析是迄今为止已知最有效的攻击迭代密码的方法之一，它利用高概率特征或差分恢复密钥。其基本思想为：通过分析明文对的差值对密文对的差值的影响来恢复某些密钥比特。简单地说，随机选取具有固定差分的一对明文，只要它们符合特定的差分条件，甚至可以不必知道它们的值；然后按照不同的概率，将输出密文中的差分分配给不同的密钥；随着对密文对的分析越来越多，将使最可能的一个密钥显现出来，这样就得到了正确的密钥。

差分密码分析最初是针对 DES 加密提出的一种攻击方法，可用于 6 轮以上的 DES 加密。8 轮 DES 需要 2^{14} 个选择明文，10 轮和 14 轮 DES 分别需要 2^{24} 和 2^{39} 个选择明文才能破解。虽然差分密码分析未能破解 16 轮的 DES 加密，但用它破解轮数较低的 DES 还是很成功的。例如，在个人计算机上几分钟就可以破解 8 轮 DES。差分密码分析除了用来攻击 DES 外，也可以被用来攻击其他密码体制。

3. 线性密码分析

线性密码分析本质上是一种已知明文攻击法，是对 DES 加密方法进行破译的主要方法。这种方法用 221 个已知明文可以破译 8 轮 DES，用 2^{47} 个明文可以破译 16 轮 DES。在某些情况下，这种方法可用于唯密文攻击。其基本思想是：通过寻找一个给定密码算法的有效的线性近似表达式来破译密码系统。由于每个密码系统均为非线性系统，因此只能寻找线性近似表达式。如果分别将明文的一些位、密文的一些位进行异或运算，然后再将这两个结果进行异或运算，这两个结果的运算结果是一个位，这一位与密钥的一些位进行异或运算的结果相同。这一位就是概率为 P 的线性近似值，在 P 不等于 1/2 的前提下，就可以使用该偏差，用得到的明文及相对应的密文便可猜测密钥的位值。得到的明文数据越多，猜测密钥的位置越可靠。概率 P 越大，用同样数据量分析的成功率就越高。

4. 差分-线性密码分析

强力攻击、差分密码分析和线性密码分析是 3 种对 DES 主要的攻击方法。由于差分密码分析和线性密码分析对于 16 轮的 DES 的分析所需的选择（已知）明文个数太大，所以目前最有效的攻击仍然是强力攻击。而差分-线性密码分析就是对差分密码分析和线性密码分析进行改进，是降低它们复杂度的众多改进之一，它利用的是差分密码分析和线性密码分析相结合的技术。

5. 插值攻击

插值攻击仅对某些密码算法有效，即轮数很少或轮函数的次数很低的算法。如果密文可以表示成明文的多项式，则插值攻击根据具体条件可以给出等价于加密或解密算法的一个变换，或者恢复出最后一轮的子密钥。该方法利用了拉格朗日插值公式的思想。插值攻击由 Knudsen 和 Jakobsen 提出：如果一个密码算法是固定的密钥的低次多项式函数，或项数较少的多项式，其项数可以估算出来，则通过插值法可以得到其代数表达式，从而可能恢复出密钥；在改进后的插值攻击中，可以精确计算出多项式函数的某些项的系数，从而在利用有限域上的傅里叶变换的基础之上，也可以求出相应的密钥。另外，如果密文可以作为两个多项式的商，且可以估计出来这两个多项式的项数，那么相应的密钥同样可以恢复出。

插值攻击使用代数函数来代表 S 盒，可以用已知明文攻击法取得此函数的样本点，再用拉格朗日插值法产生。这个代数函数可能是在有限体上的有理函数、多项式函数或二次函数。此函数也可以用选择明文攻击法取得样本点，这样可以简化所使用的代数函数，让攻击效率更高。Thoms Jakobsen 又将概率的概念引入了插值攻击法，通过 MadhuSudan 演算法来改善其对 RS(Reed-Solomon)纠错码的解译能力。如此一来在明文与密文的内容仅有极少的代数关系时插值攻击也有效。

3.4 非对称密钥密码技术

3.4.1 基本概念

非对称密钥密码算法（即公开密钥算法）的思想最早是由当时在美国斯坦福大学的 Diffie 和 Hellman 两人于 1976 年在其论文 *New Direction in Cryptography* 中提出的。但

目前最流行的 RSA 算法是 1977 年由 MIT 教授 Ronald L. Rivest、Adi Shamir 和 Leonard M. Adleman 共同开发的，分别取自 3 名数学家的名字的第一个字母来构成的。

1976 年提出的公开密钥密码体制思想不同于传统的对称密钥密码体制，它要求密钥成对出现，一个为加密密钥 e，另一个为解密密钥 d，且不可能从其中一个推导出另一个。自 1976 年以来，已经提出了多种公开密钥密码算法，其中许多是不安全的，一些认为是安全的算法又有许多是不实用的，它们要么是密钥太大，要么密文扩展十分严重。多数密码算法的安全基础是基于一些数学难题，这些难题专家们认为在短期内不可能得到解决，因为一些问题（例如因子分解问题）至今已有数千年的历史了。

非对称加密算法使用两对密钥：一个公共密钥和一个专用密钥。用户要保障专用密钥的安全，而公共密钥则可以发布出去。公共密钥与专用密钥是有紧密关系的，用公共密钥加密的信息只能用于专用密钥解密，反之亦然。由于公钥算法不需要联机密钥服务器，密钥分配协议简单，所以极大简化了密钥管理。除加密功能外，公钥系统还可以提供数字签名。非对称密码算法解决了对称密码体制中密钥管理的难题，并提供了对信息发送人的身份进行验证的手段，是现代密码学的最重要的发明和进展。

单向函数和陷门单向函数的概念是公钥密码学的核心，可以说非对称密钥密码体制的设计就是陷门单向函数的设计。

给定任意两个集合 X 和 Y。函数 $f: X \rightarrow Y$ 称为单向的，如果对每一个 x 属于 X，很容易计算出函数 $f(x)$的值，而对大多数 y 属于 Y，要确定满足 $y=f(x)$的 x 是计算上困难的（假设至少有这样一个 x 存在）。注意，不能将单向函数的概念与数学意义上的不可逆函数的概念混同，因为单向函数可能是一个数学意义上可逆或者一对一的函数，而一个不可逆函数却不一定是单向函数。

目前，还没有人能够从理论上证明单向函数是存在的。单向函数存在性的证明将意味着计算机科学中一个最具挑战性的猜想 $P=NP$，即 NP 完全问题的解决，而关于 NP 完全性的理论却不足以证明单向函数的存在。现实中却存在几个单向函数的“候选”。说他们是“候选”，是因为他们表现出了单向函数的性质，但还没有办法从理论上证明它们一定是单向函数。

显然，单向函数不能直接用作密码体制，因为如果用单向函数对明文进行加密，即使是合法的接收者也不能还原出明文了，因为单向函数的逆运算是困难的。与密码体制关系更为密切的概念是陷门单向函数。一个函数 $f: X \rightarrow Y$ 称为是陷门单向的，如果该函数及其逆函数的计算都存在有效的算法，而且可以将计算 f 的方法公开，即使由计算 f 的完整方法也不能推导出其逆运算的有效算法。其中，使得双向都能有效计算的秘密信息叫做陷门（trap door）。

需要注意的是，不能顾名思义地认为陷门单向函数是单向函数。事实上，陷门单向函数不是单向函数，它只是对于那些不知道陷门的人表现出了单向函数的特性。

3.4.2　RSA 算法

RSA 密码体制是目前为止最为成功的非对称密码算法，它的安全性是建立在“大数分解和素性检测”这个数论难题的基础上的，即将两个大素数相乘在计算上容易实现，而将该乘积分解为两个大素数因子的计算量相当大。虽然它的安全性还未能得到理论证明，但经过 20 多年的密码分析和攻击，迄今仍然被实践证明是安全的。

RSA 使用两个密钥，一个是公共密钥，一个是私有密钥。若用其中一个加密，则可用另一个解密，密钥长度从 40 到 2048bit 可变，加密时也把明文分成块，块的大小可变，但不能超过密钥的长度。RSA 算法把每一块明文转化为与密钥长度相同的密文块。密钥越长，加密效果越好，但加密/解密的开销也大，所以要在安全与性能之间折中考虑，一般 64 位是较合适的。RSA 的一个比较知名的应用是 SSL，在美国和加拿大 SSL 用 128 位 RSA 算法，由于出口限制，在其他地区通用的则是 40 位版本。

RSA 算法研制的最初理念与目标是努力使 Internet 安全可靠，旨在解决 DES 算法秘密密钥利用公开信道传输分发的难题。而实际结果不但很好地解决了这个难题，还可利用 RSA 来完成对电文的数字签名以对抗电文的否认与抵赖，同时还可以利用数字签名较容易地发现攻击者对电文的非法篡改，以保护数据信息的完整性。

RSA 算法描述如下：

1. 密钥生成

选择两个互异的大素数 p 和 q，n 是二者的乘积，即 $n = pq$，使 $\Phi(n)=(p-1)(q-1)$，$\Phi(n)$为欧拉函数。随机选取正整数 e，使其满足 $\gcd(e,\Phi(n))=1$，即 e 和 $\Phi(n)$互质，则将(n,e)作为公钥。

求出正数 d，使其满足 $e\times d=1 \bmod \Phi(n)$，则将$(n,d)$作为私钥。

2. 加密算法

对于明文 M，由 $C=M^e \bmod n$，得到密文 C。

3. 解密算法

对于密文 C，由 $M=C^d \bmod n$，得到明文 M。

如果窃密者获得了 n、e 和密文 C，为了破解密文他必须计算出私钥 d，为此需要先分解 n 为 p 和 q。为了提高破解难度，达到更高的安全性，一般商业应用要求 n 的长度不小于 1024bit，更重要的场合不小于 2048bit。

RSA 算法提出以后，引起了许多密码分析学家的兴趣，提出了一些针对于 RSA 的攻击方法，例如，对 RSA 的公共模数攻击；对 RSA 的低加密指数攻击；对 RSA 的低解密指数攻击；对 RSA 的选择密文攻击等。根据这些成功的攻击，Jadith Moore 列出了使用 RSA 的一些限制：

(1) 知道对于一个给定模数的一个加/解密密钥指数对，攻击者就能够分解这个模数，无须分解 n 就可以计算出别的加/解密对。

(2) 在通信网络中，利用 RSA 的协议不应该使用公共模数。

(3) 消息中应该使用随机序列填充以避免对加密指数的攻击。

(4) 解密指数应该大。

属于基于大整数因式分解困难问题的公钥密码体系的公钥密码还包括 Rabin 算法和 Williams 算法，这里不作详细介绍。

例 3-9 选择两个大素数 $p=7$、$q=17$，$p\neq q$

计算：$n=pq=7\times17=119$，$\Phi(n)=(p-1)(q-1)=6\times16=96$。96 的因子有 2、3，因此 e 不能有 2 和 3 的因子；

选择整数 $e=5$(公钥，即加密密钥)，使 $\gcd(e,\Phi(n))=1$；

选择整数 $d=77$(私钥,即解密密钥),使 $d\times e \bmod \Phi(n)=1$,$(5\times 77) \bmod 96=385 \bmod 96=1$

公钥:$K_U=\{e,n\}=\{5,119\}$;

私钥:$K_R=\{d,n\}=\{77,119\}$;

再加密:$C=M^e \bmod n$;

解密:$M=C^d \bmod n$。

3.4.3　ElGamal 算法

ElGamal 为目前著名的公开密钥密码系统之一,是由 ElGamal 于 1985 年提出的。ElGamal 密码系统可作为加/解密、数字签名等之用,其安全性是建立于离散对数(Discrete Logarithm)问题,即给定 g、p 与 y,求 x 为计算上不可行。ElGamal 算法包括密钥生成、加密过程、解密过程。

1. 密钥生成

(1) 任选一个大质数 p,使得 $p-1$ 有大质因数。

(2) 任选一个 mod p 之原根 g。

(3) 公布 p 与 g。

使用者任选一私钥 $x\in Z_p$,并计算密钥 $y=g^x \bmod p$。

2. 加密过程

(1) 任选一个数 $r\in Z_p$ 满足 $\gcd(r,p-1)=1$,并计算:

$$c_1 = g^r \bmod p, c_2 = m \times y^r \bmod p$$

(2) 密文为$\{c_1,c_2\}$。

3. 解密过程

(1) 计算 $w=(c_1^x)^{-1} \bmod p$。

(2) 计算明文 $m=c_2\times w \bmod p$。

ElGamal 方法具有以下优点:

(1) 系统不需要保存秘密参数,所有的系统参数均可公开。

(2) 同一个明文在不同的时间由相同加密者加密会产生不同的密文(概率式密码系统),但 ElGamal 方法的计算复杂度比 RSA 方法要大。

例 3-10　$a=\{2,3,6,13,27\}$,$m=53$,选一个与 m 互质的数 $w=13$,已知明文 $p=10101$。

$a'=(w\times a) \bmod m$,则密文 $c=\sum a'_i\times x_i \bmod m=(26+39+33) \bmod 53=31$。

$w^{-1}=49$,$s'=w^{-1}s \bmod m=49\times 31 \bmod 53=35$,用私钥解密可得:$35=2+6+27$,故 $p=10101$

3.4.4　椭圆曲线算法

公开密钥密码学的数学理论早在百年前就已经很完备了,RSA、ElGamal 等密码系统都是如此,而椭圆曲线在代数学与几何学上广泛的研究已超出百年之久,已有丰富且深入的理论,而椭圆曲线系统第一次应用于密码学上是于 1985 年由 Koblitz 与 Miller 分别提出的,

随后有两个较著名的椭圆曲线密码系统被提出：一种是利用 ElGamal 的加密法；另一种为 Menezes-Vanstone 的加密法。以下将介绍椭圆曲线的定义、加法运算和反元素运算。

1. 椭圆曲线定义

令 $p>3$ 为质数，在 GF(p)中的椭圆曲线 $E: y^2=x^3+ax+b \bmod p$，其中：$4a^3+27b^2 \neq 0 (\bmod p)$。曲线上另定义一个无穷远点 O，对任一点 $A \in E$，$A+O=O+A=A$。

2. 加法运算

令 $A=(x_1,y_1)$ 与 $B=(x_2,y_2)$为 E 上的点，若 $x_2=x_1$ 且 $y_2=-y_1$，则 $A+B=O$；否则 $A+B=(x_3,y_3)$，其中

$$\begin{cases} x_3 = \lambda^2 - x_1 - x_2 \\ y_3 = \lambda(x_1 - x_3) - y_1 \end{cases} \tag{3-23}$$

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, & A \neq B \\ \dfrac{3x_1^2 + a}{2y_1}, & A = B \end{cases} \tag{3-24}$$

注意：椭圆曲线运算中，大写参数表示点，小写参数表示数值。椭圆曲线中的乘法运算是透过加法运算达成的。为了加快速度，可以用倍加的运算来达成。例如 $4P$ 计算时，由于 $4P=2P+2P$，再计算 $2P=P+P$ 即可。

3. 反元素运算

点 $A=(x,y)$的反元素为 $-A=-(x,y)=(x,-y)$。$A+(-A)=(-A)+A=O$，此时 O 称为乘法单位元素。

例 3-11 在椭圆曲线 $E: y^2=x^3+x+6 (\bmod 11)$上的点有：

(2,4)，(2,7)，(3,5)，(3,6)，(5,2)，(5,9)，(7,2)，(7,9)，(8,3)，(8,8)，(10,2)，(10,9)。再加上 O 共有 13 点。注意在计算点时，要检验 x^3+x+6 之值是否属于 QR_{11}。除了 O 以外，任意点均可以视为 E 的始元素(Primitive Element)。

注意：令定义于 Z_p 的椭圆曲线 E 的所有点的个数为 $\# E$，则满足：

$$p+1-2\sqrt{p} \leqslant \# E \leqslant p+1+2\sqrt{p} \tag{3-25}$$

4. 椭圆曲线密码体制

设 GF(p)是一个有限域，GF(p)上的椭圆曲线是指满足 Weirstrass 方程：

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_5 \text{(无奇点，} a_1, a_2, a_3, a_4, a_5 \in \text{GF}(p)\text{)} \tag{3-26}$$

的所有解(x,y)与无穷远点 O 构成的非空集合。

选取一点 $G \in E(\text{GF}(p))$作为公共基点，要求这个公共基点的阶 $l=\text{ord } G$ 是一个素数阶，G 为生成元，$\langle G \rangle$是由点 G 生成的 p 阶循环子群。

对于 $Q=dG$，$d \in Z_q$，G、$Q \in E(\text{GF}(p))$，已知 G、Q 求 d 称为 ECDLP(椭圆曲线离散对数问题)。基于椭圆曲线的各种密码体制的安全性最终可归结为解 ECDLP 问题，当数据量足够大以致 ECDLP 问题无法解决时，就认为该密码体制是安全的，具有 160bit 数据长度的 ECDLP 问题在目前被认为是安全的。

一般的椭圆曲线密码体制都基于以下运算：

(1) 存在一个容易计算的函数 $f: m \to P(m)$。

(2) 选取整数 e,$1<e<N$,选取整数 d,使得 $de=1 \pmod N$,由 $deP(m)=P(m)$,可恢复 $P(m)$。

(3) 选取整数 a,$1<a<N$,由 $P(m)=P(m)+aP-aP$,可恢复出 $P(m)$。

SECG 的标准文档 SEC1 中,对有限域 GF(p)上的椭圆曲线域参数 T 定义为如下六元组:$T=(p,a,b,G,n,h)$。其中,$a,b\in \mathrm{GF}(p)$,满足方程 $y^2\equiv x^3+ax+b$,$G=(x_G,y_G)$为曲线上的基点,基点 G 的阶 n 为一素数,整数 h 为余因子,$h=\#E(\mathrm{GF}(p))/n$,$\#E(\mathrm{GF}(p))$为椭圆曲线的阶。

一个典型的椭圆曲线公钥密码可以描述如下:

设 p 是不等于 3 的素数,椭圆曲线 $E(\mathrm{GF}(p))$包含一个循环子群 A,在 A 中离散对数问题是难处理的。选取 $\alpha\in E$,$0<\alpha<\#A-1$,计算 $\beta=a\alpha$,将 α,β 值公开作为公钥,保密 a 作为私钥。

1) 加密过程

设明文 $m=(m_1,m_2)\in Z_p^*\times Z_p^*$,即 m_1、m_2 均属于 Z_p^*,对明文加密如下:

(1) 选取一整数 k,$0<k<\#A-1$,k 保密。

(2) 计算:$y_0=k\alpha$,$(c_1,c_2)=k\beta$,$y_1=c_1m_1 \pmod q$,$y_2=c_2m_2 \pmod q$。

(3) 则密文 $c=(y_0,y_1,y_2)$,将其发送给接收方。

2) 解密过程

(1) 接收方接收到密文 c。

(2) 计算:$(c_1,c_2)=ay_0$。

(3) 通过下列运算恢复明文:$m=(y_1c_1^{-1} \pmod q, y_2c_2^{-1} \pmod q)$。

椭圆曲线是一种能够适应未来通信技术和信息安全技术发展的新型密码体制。对于 q 元有限域上的椭圆曲线,q 为 160bit 时,RSA 密码体制需要 1024bit 的模数才能达到同等的安全强度。也就是说,椭圆曲线密码体制在相同的安全强度下所要求的密钥强度仅是 RSA 的 1/6,因此在运算速度和存储空间方面具有很大的优势,在实际应用中具有很大的使用价值。

3.4.5 混合加密算法

表 3-13 列出了对称密钥加密与非对称密钥加密的区别,结合二者的优点可以设计出混合加密算法。

表 3-13 对称与非对称密钥加密对比

特　性	对称密钥加密	非对称密钥加密
加密/解密使用的密钥	加密/解密使用的密钥相同	加密/解密使用的密钥不同
加密/解密的速度	快	慢
得到的密文长度	通常等于或小于明文长度	大于明文长度
所需密钥数与消息交换参与者个数的关系	大约为参与者个数的平方,因此伸缩性不好	等于参与者个数,因此伸缩性好
用法	主要用于加密/解密,不能用于数字签名	可以用于加密/解密和数字签名(完整性和不可抵赖性)

由前文的介绍可知，对称密钥密码体制中的DES算法具有可靠性较高(16轮变换，增大了混乱性和扩散性，输出不残存统计信息)、加密/解密速度快、算法容易实现(可用软件和硬件实现，硬件实现速度快)以及通用性强等优点，但也存在密钥位数少、弱密钥和半弱密钥、易于遭受穷尽攻击以及密钥管理复杂等缺点。与DES算法相比，RSA算法具有以下优点：

(1) 密钥空间大。

(2) 密钥管理简单，网上每个用户仅保密一个密钥，不需要密钥配送。

(3) 便于数字签名。

(4) 可靠性较高，取决于分解大素数的难易程度。

RSA算法的缺点是加密/解密速度慢、算法复杂。如果RSA和DES结合使用，则正好弥补RSA的缺点。即DES用于明文加密，RSA用于DES密钥的加密。由于DES加密速度快，适合加密较长的报文，而RSA可解决DES密钥分配的问题。一种混合了非对称和对称加密算法的加密方式如图3-11所示。

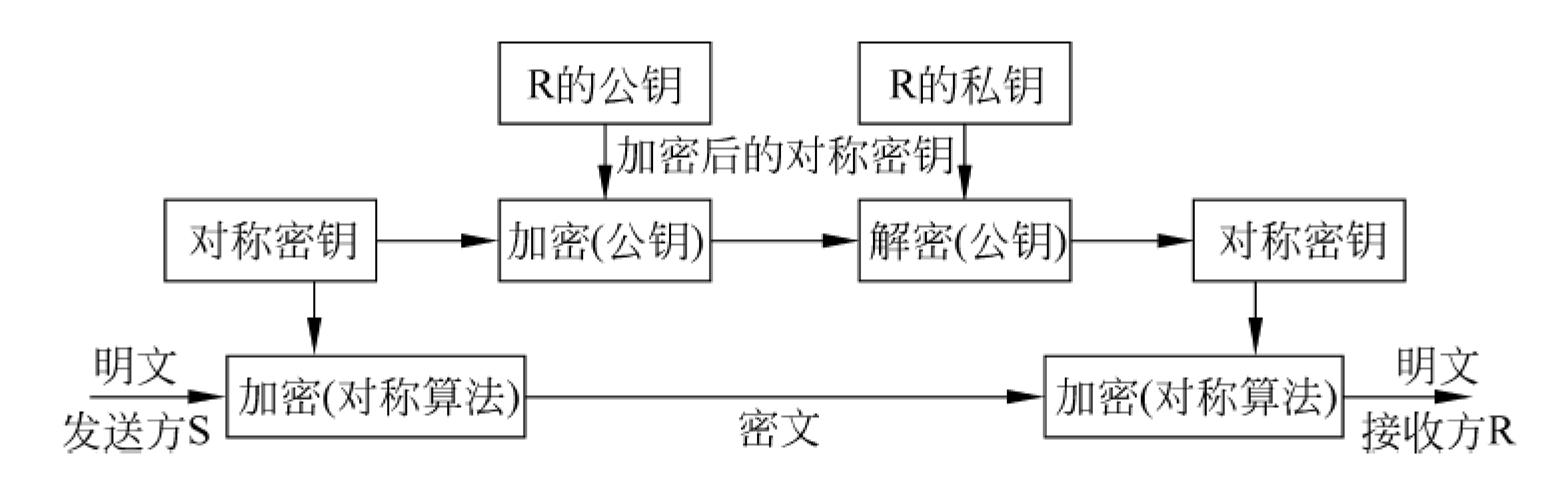


图3-11 混合加密方式

这种混合加密方式的原理是：发送方S先使用DES或IDEA对称算法对数据进行加密，然后使用公钥算法RSA加密前者的对称密钥；接收方R先使用RSA算法解密出对称密钥，再用对称密钥解密被加密的数据。要加密的数据量通常很大，但因对称算法对每个分组的处理只需很短的时间便可完成，因此对大量数据的加密/解密不会影响效率。

实际网络多采用双钥和单钥密码相结合的混合加密体制，即加/解密时采用单钥密码，密钥传输则采用双钥密码。这样既解决了密钥管理的困难，又解决了加密和解密速度的问题。

3.5 信息认证技术概述

认证(Authentication)又称为鉴别，它是证实某事物是否名副其实或是否有效的一个过程。认证是防止主动攻击(例如篡改、伪造信息等)的一项重要技术，可用于开放环境中各种信息系统安全性的保护。它采用一定的技术方式解决网络数据传输过程中可能出现的非法访问与篡改、假冒伪造、拒绝服务、抵赖等安全问题，确保网络中传输数据的机密性、访问可控制性、数据完整性、抗抵赖性等方面的安全需求。认证的目的包括以下两个方面：

(1) 消息完整性认证：即验证数据在传输或存储过程中是否被篡改、重放或延迟等。

(2) 身份认证：即验证信息发送者和接收者的身份是否合法，可以通过密码或密钥的

方式确认。

认证与加密的主要区别在于：加密用于确保数据的保密性，防止非法者的被动攻击，例如截取、窃听等；而认证多用于确保报文发送者和接收者的真实性以及报文的完整性，阻止对手的主动攻击，例如冒充、篡改等。认证是许多应用系统中安全保护的第一道防线，因此对于用户来说是非常重要的。

一个安全的认证体制至少应该满足以下要求：

(1) 接收者能够检验和证实消息的合法性、真实性和完整性。

(2) 消息的发送者对所发的消息不能抵赖，某些场合也要求消息的接收者不能否认收到的消息。

(3) 除了合法的消息发送者外，其他人不能伪造发送消息。

通常，认证和保密的关系是相对独立的，即一个认证系统不能自动地提供保密的功能，而一个保密系统也不会自然地提供认证的功能。图 3-12 给出的是一个纯认证系统的模型。

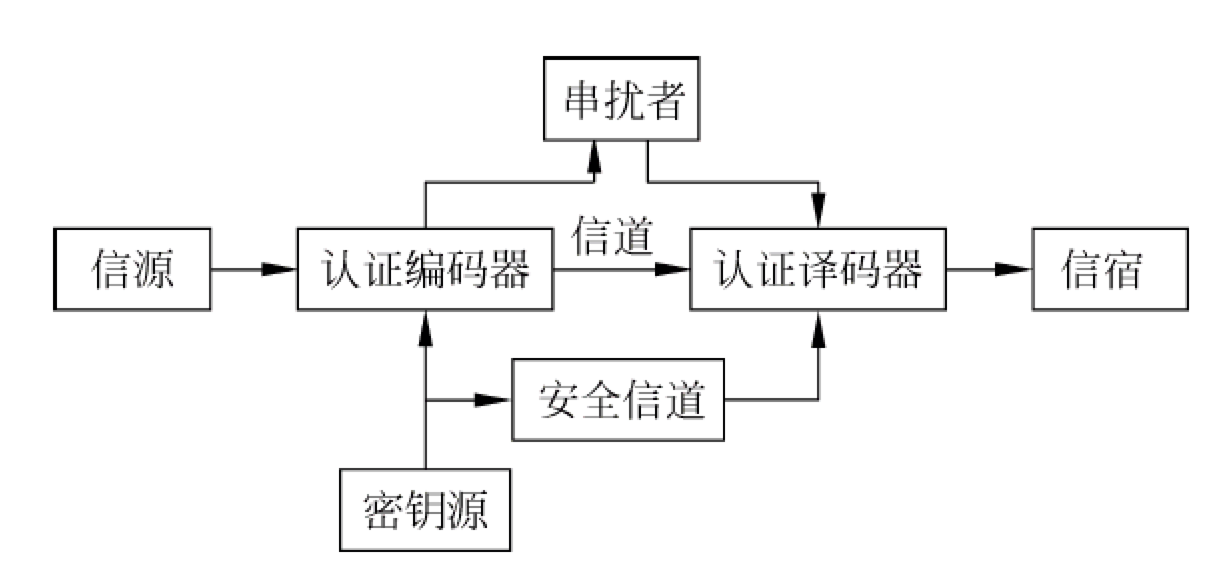


图 3-12　一个纯认证系统的模型

在这个系统中发送者(信源)通过一个公开信道将信息传输给接收者(信宿)，接收者除了收到消息本身以外，还要通过认证编码器和认证译码器验证消息是否被篡改以及消息是否来自合法的发送者。系统的串扰者是指可截获和分析信道中传输的密文，而且可伪造密文送给接收者进行欺诈的主动攻击者。

信息认证是指通过对消息或者消息有关的信息进行加密或签名变换进行的认证，目的是为了防止传输和存储的消息被有意/无意的篡改，包括消息内容认证(即消息完整性认证)、消息的源和宿认证(即身份认证)，以及消息的序号和操作时间认证等。它在票据防伪中具有重要应用(例如税务的金税系统和银行的支付密码器)。信息认证主要用于防止信息被篡改。

3.6　Hash 函数与消息认证

3.6.1　基本概念

1. Hash 函数

Hash 函数长期以来一直在计算机科学中使用，无论从数学角度或别的角度看，Hash 函数就是把可变长度的输入串转换成固定长度的(经常更短)输出串(叫做 Hash 值)的一种函数。

Hash 函数具备以下性质：

(1) Hash 函数 H 可适用于任意长度的输入数据块，产生固定长度的 Hash 值。

(2) 对于每一个给定输入数据 M，都能很容易计算出它的 Hash 值 $H(M)$。

(3) 如果给定 Hash 值 h，要逆向推出输入数据 M 在计算上不可行，即 Hash 函数具备单向性。

(4) 对于给定的消息 M_1 和其 Hash 值 $H(M_1)$，找到满足 $M_2 \neq M_1$，且 $H(M_2)=H(M_1)$ 的 M_2 在计算上是不可行的，即抗弱碰撞性。

(5) 要找到任何满足 $H(M_1)=H(M_2)$ 且 $M_1 \neq M_2$ 的消息对 (M_1, M_2) 在计算上是不可行的，即抗强碰撞性。

这里所说的碰撞(Collision)，是指如果有两个不同的消息，它们生成的 Hash 值相同，则称发生了一次碰撞。特别需要注意的是，Hash 函数并不提供机密性，且无须使用密钥就可以生成 Hash 值，它非常适合于消息认证。

安全单向 Hash 函数的一般结构如图 3-13 所示。

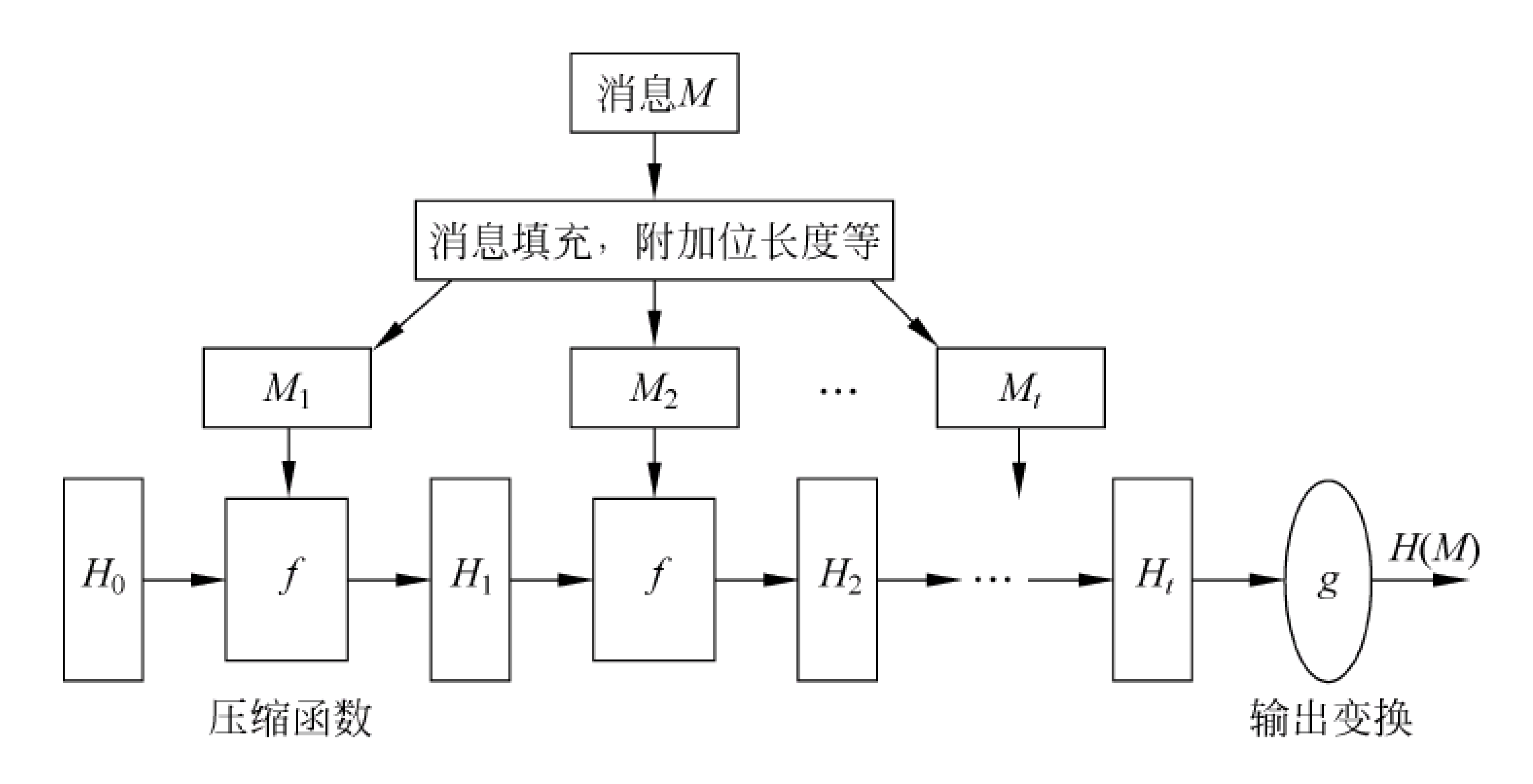


图 3-13　安全单向 Hash 函数的一般结构

由图 3-13 可知，单向 Hash 函数重复使用一个压缩函数 f 来实现 Hash 值的生成。压缩函数通常有两个输入，一个是前一阶段的输出 H_{i-1}，另一个来源于消息分组 M_i，最后产生一个输出 H_i，可表达为：

$$M_i = f(H_{i-1}, M_i) \quad (i = 1, 2, \cdots, t) \tag{3-27}$$

通常 H_0 为初始向量。

2. 消息认证

消息鉴别码(Message Authentication Code，MAC)也叫密码校验和(Cryptographic Checksum)，是鉴别函数的一种。

消息鉴别码鉴别的原理是：用公开函数和密钥产生一个固定长度的值作为认证标识，用这个标识鉴别消息的完整性。使用一个密钥生成一个固定大小的小数据块，即 MAC，并将其加入到消息中，然后传输。接收方利用与发送方共享的密钥进行鉴别认证等。

消息认证码是利用密钥对要认证的消息产生新的数据块并对数据块加密生成的，它对于要保护的信息来说是唯一的和一一对应的。因此可以有效地保护消息的完整性，以及实现发送方消息的不可抵赖和不能伪造。消息认证码的安全性取决于两点：一是采用的加密

算法；二是待加密数据块的生成方法。

消息认证不支持可逆性，是多对一的函数，其定义域由任意长的消息组成，而值域则是由远小于消息长度的比特构成。从理论上说，一定存在不同的消息产生相同的冗余数据块。因此必须要找到一种足够单向和强碰撞自由性的方法对消息认证才是安全的。

首先，利用校验码加密的方式构造认证码，它可以实现数据完整性，它对消息不可抵赖、不可伪造性的认证性能取决于加密的函数。因此这种方法的安全性取决于校验码的长度和加密的方法。但是由于它是针对局部变量的校验，例如针对一行或者一列，它的抗碰撞性能不是很好，即有可能产生消息被改动而认证码却不变的情况。

其次，对于用单向 Hash 函数构造认证码的方式来说，安全性是基于该函数的抗强碰撞性的，即攻击主要目标时找到一对或更多对碰撞消息，该消息生成的摘要是相同的。在目前已有的攻击方案中，一些是一般的方法，是基于穷举的，可攻击任何类型的 Hash 方案，例如生日攻击方法；另一些是特殊的方法，只能用于攻击某些特殊类型的 Hash 方案，例如适用于攻击具有分组链结构的 Hash 方案的中间相遇攻击，适用于攻击基于模算术的 Hash 函数的修正分组攻击。因此摘要的长度是一个关键的因素。

3.6.2　常见的单向 Hash 函数

常见的单向 Hash 函数包括 MD5、SHA-1、Tiger hash 和 CRC 等。

1. MD5

MD5(Message Digest 5)是 RSA 数据安全公司开发的一种单向 Hash 算法。MD5 被广泛使用，可以用来把不同长度的数据块进行运算处理生成一个 128bit 的数据块。

MD5 算法可简要地叙述为：MD5 以 512bit 分组来处理输入的信息，且每一分组又被划分为 16 个 32bit 的子分组，经过了一系列的处理后，算法的输出由 4 个 32bit 分组组成，最终将这 4 个 32 位分组级联后将生成一个 128 位 Hash 值。MD5 算法的总体框架如图 3-14 所示。

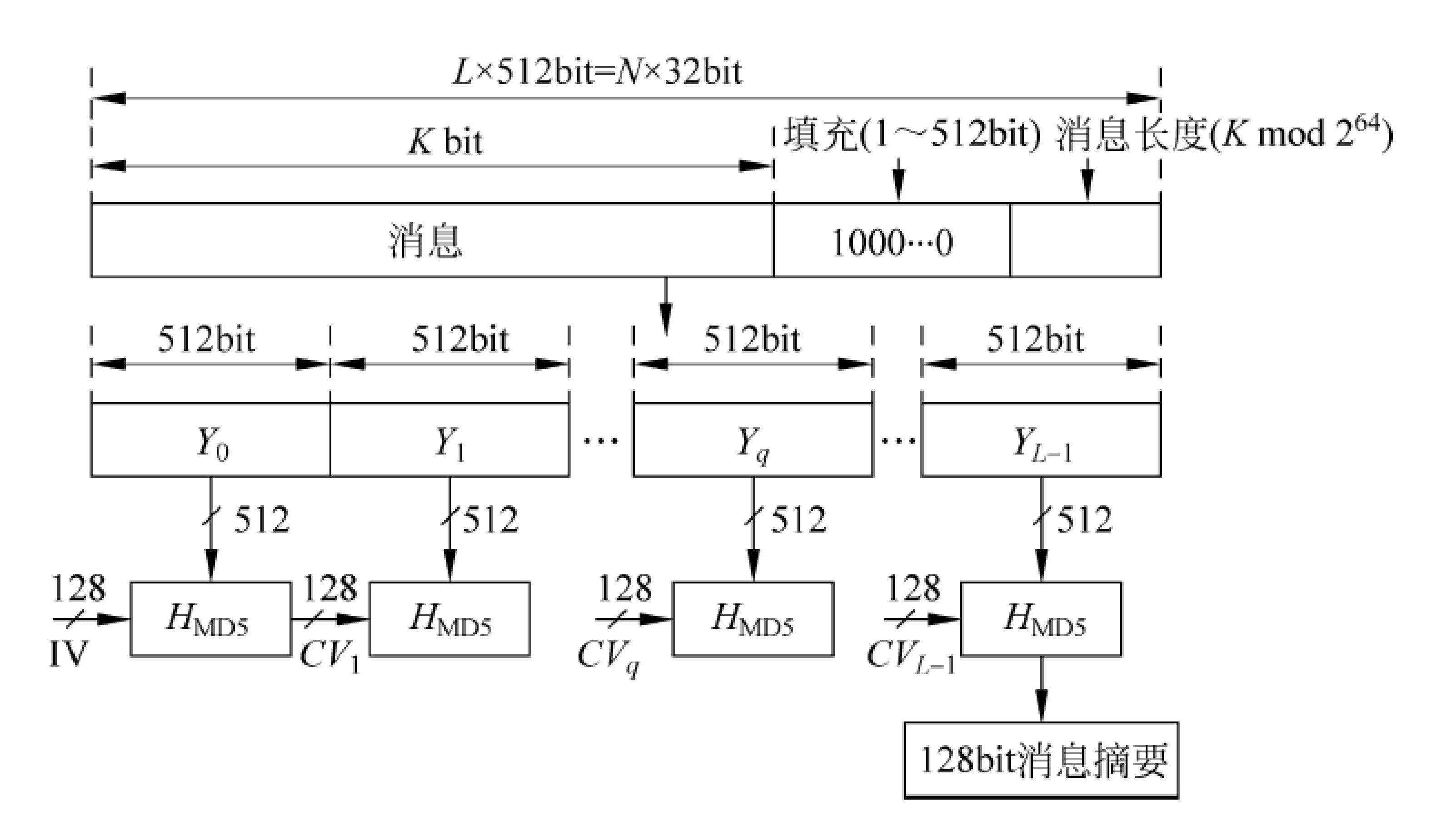


图 3-14　MD5 算法的总体框架图

在 MD5 算法中，首先需要对信息进行填充，使其位长度满足模 512 等于 448。因此，信息的位长度将被扩展至 $N\times512+448$，即 $N\times64+56$ 个字节，N 为一个非负整数。填充的

方法如下：在信息的后面填充一个1和无数个0，直到满足上面的条件时才停止用0对信息的填充；然后，再在这个结果后面附加一个以64位二进制表示的填充前信息长度。经过这两步的处理，现在的信息字节长度$=N\times512+448+64=(N+1)\times512$，即长度恰好是512的整数倍。这样做的原因是为满足后面处理中对信息长度的要求。对于单个信息分组，其处理过程如图3-15所示。

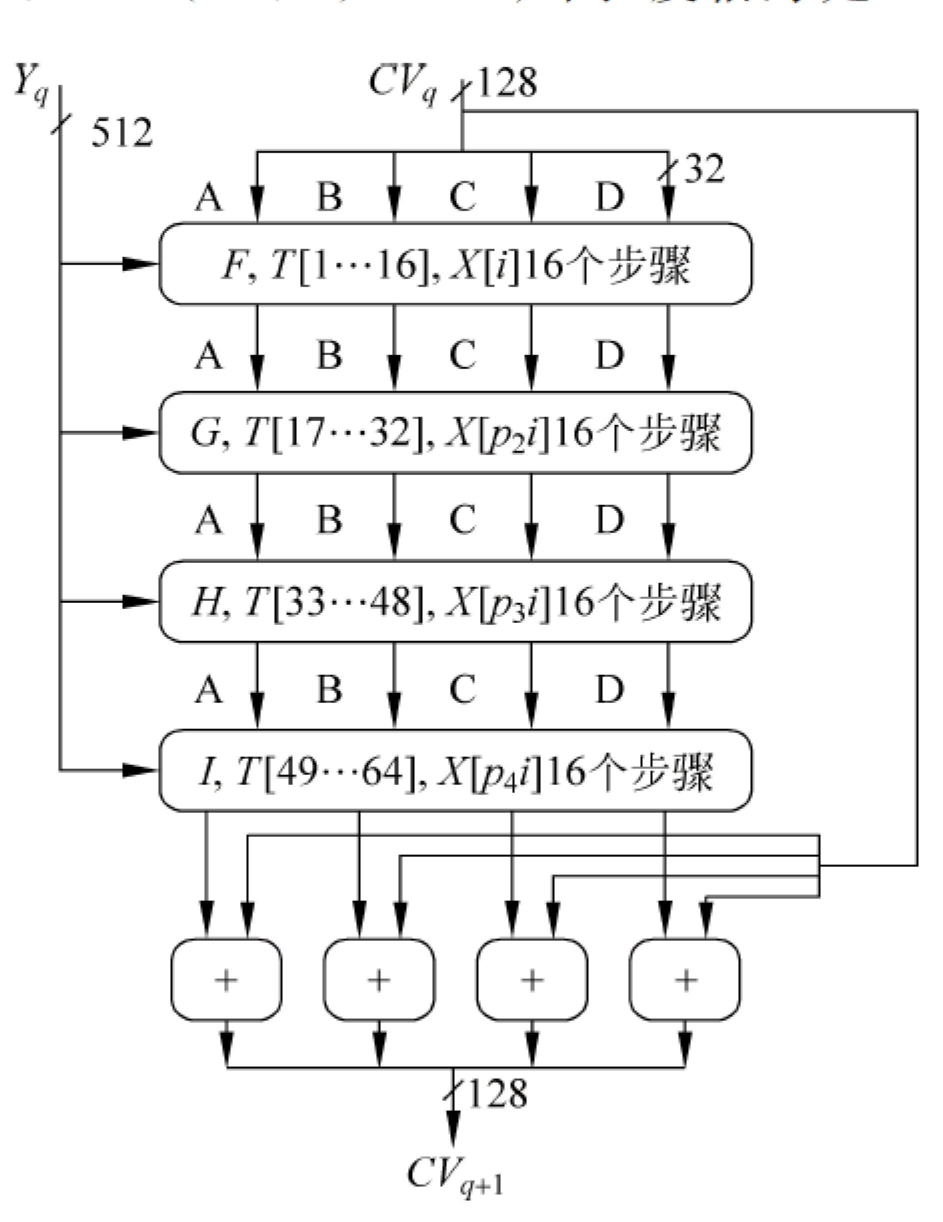


图3-15 MD5对单个512位分组的处理过程

MD5中有4个32位被称作链接变量(Chaining Variable)的整数参数，分别为：A=0x01234567，B=0x89abcdef，C=0xfedcba98，D=0x76543210。当设置好这4个链接变量后，就开始进入算法的四轮循环运算。循环的次数是信息中512位信息分组的数目。

主循环有四轮(MD4只有三轮)，每轮循环都很相似。第一轮进行16次操作。每次操作对A、B、C和D中的其中3个作一次非线性函数运算，然后将所得结果加上第4个变量、文本的一个子分组和一个常数。再将所得结果向右环移一个不定的数，并加上A、B、C或D中之一。最后用该结果取代A、B、C或D中之一。

以下是每次操作中用到的4个非线性函数(每轮一个)：

(1) $F(X,Y,Z)=(X\&Y)\mid(\overline{X}\&Z)$。

(2) $G(X,Y,Z)=(X\&Z)\mid(Y\&\overline{Z})$。

(3) $H(X,Y,Z)=X\oplus Y\oplus Z$。

(4) $I(X,Y,Z)=Y\oplus(X\mid(\overline{Z}))$。

其中，&是与，|是或，$^-$是非，$\oplus$是异或。

这4个函数的说明：如果X、Y和Z的对应位是独立和均匀的，那么结果的每一位也应是独立和均匀的。F是一个逐位运算的函数，即：如果X，那么Y，否则Z。函数H是逐位奇偶操作符。

每一轮都会使用到一个64元素表$T[1\cdots64]$中的四分之一，$T[1\cdots64]$表是通过正弦函数构造得到的。T中的第i个元素表示为$T[i]$，它等于$2^{32}\times\text{abs}(\sin(i))$的整数部分值，$i$的单位是弧度。

在MD5算法中，其核心是压缩函数H_{MD5}。MD5的压缩函数中有4次循环，每一次循环包含对缓冲区A、B、C、D的16步操作，每一循环的形式为：

$$(a,b,c,d)=(d,b+((a+g(b,c,d)+X[k]+T[i])<<s),b,c)$$

其中，a、b、c、d对应着缓冲区A、B、C、D中的4个字；g表示F、G、H、I中的某一个函数；$X[k]$表示当前512位数据块Y_q中的第k个32位；$<<s$表示把32位循环左移s位；+是mod 2^{32}。MD5的基本操作如图3-16所示。

2. SHA-1

安全Hash算法(SHA)是由美国NIST开发的，作为联邦信息处理标准FIPS PUB 180

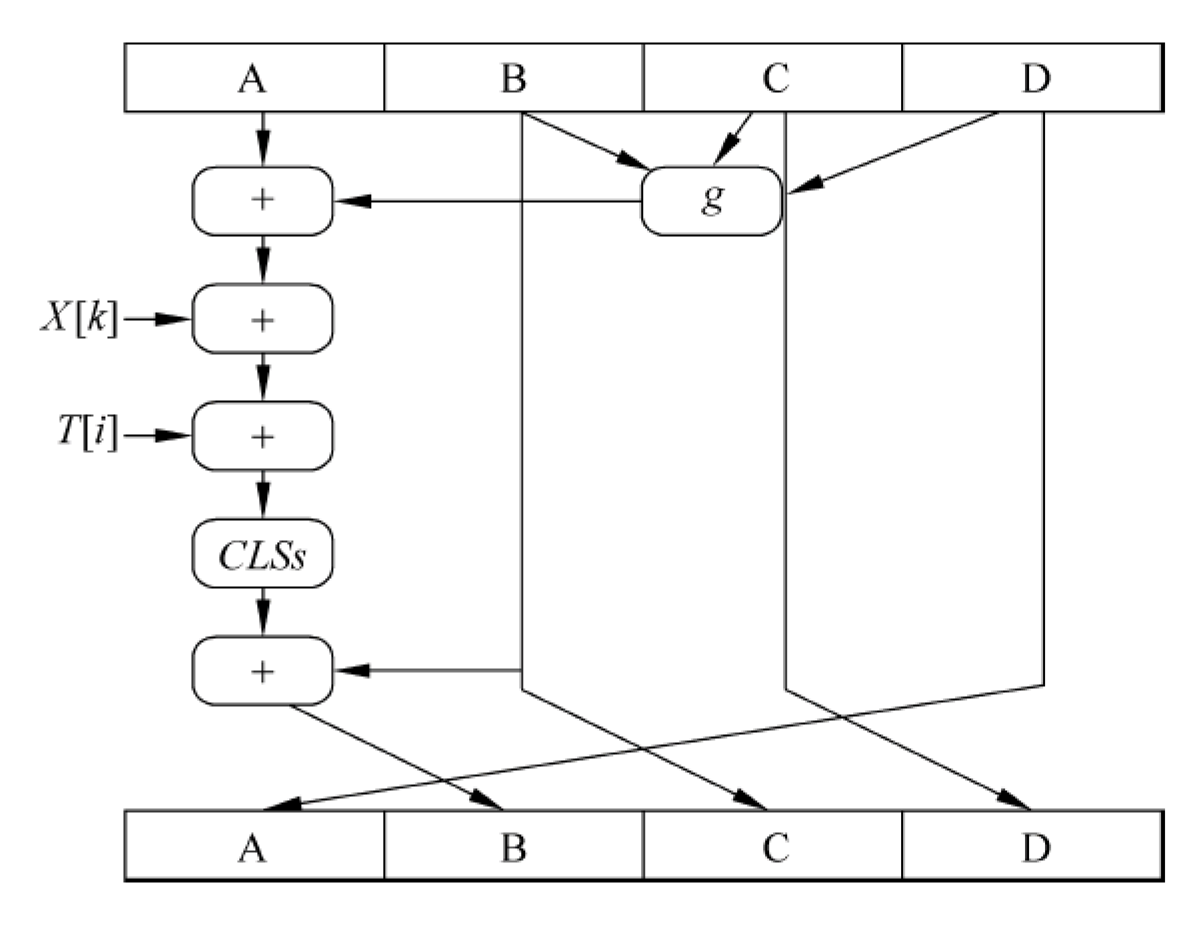


图 3-16　基本的 MD5 操作

于 1993 年发表，在 1995 年修订以后，被称为 SHA-1(即 FIPS PUB 180-1 标准)。SHA-1 是基于 MD4 算法设计的。

SHA-1 主要适用于数字签名标准(Digital Signature Standard，DSS)里面定义的数字签名算法(Digital Signature Algorithm，DSA)。对于长度小于 2^{64} 位的消息，SHA-1 会产生一个 160 位的消息摘要。当接收到消息的时候，这个消息摘要可以用来验证数据的完整性。在传输的过程中，数据很可能会发生变化，那么这时候就会产生不同的消息摘要。SHA-1 有如下特性：不可以从消息摘要中复原信息；两个不同的消息不会产生同样的消息摘要。

在 SHA-1 算法处理步骤包括如下 5 步：

(1) 添加填充位。SHA-1 算法对信息的填充和 MD5 采用的办法完全一样。

(2) 添加长度。一个 64 位的数据块，表示原始消息的长度。

(3) 初始化消息摘要的缓冲区(即 IV 值)。消息缓冲区包括 160 位，用 5 个 32 位的寄存器(A、B、C、D、E)表示，用来存储中间和最终 Hash 函数的结果。用 a、b、c、d、e 分别表示这 5 个寄存器中的字，初始化为(十六进制表示)：

a=67452301；

b=EFCDAB89；

c=98BADCFE；

d=10325476；

e=C3D2E1F0。

(4) 以 512 位数据块作为单位来对消息进行处理。算法的核心是一个包含 4 个循环的模块，每个循环由 20 个处理步骤组成，其处理过程如图 3-17 所示。

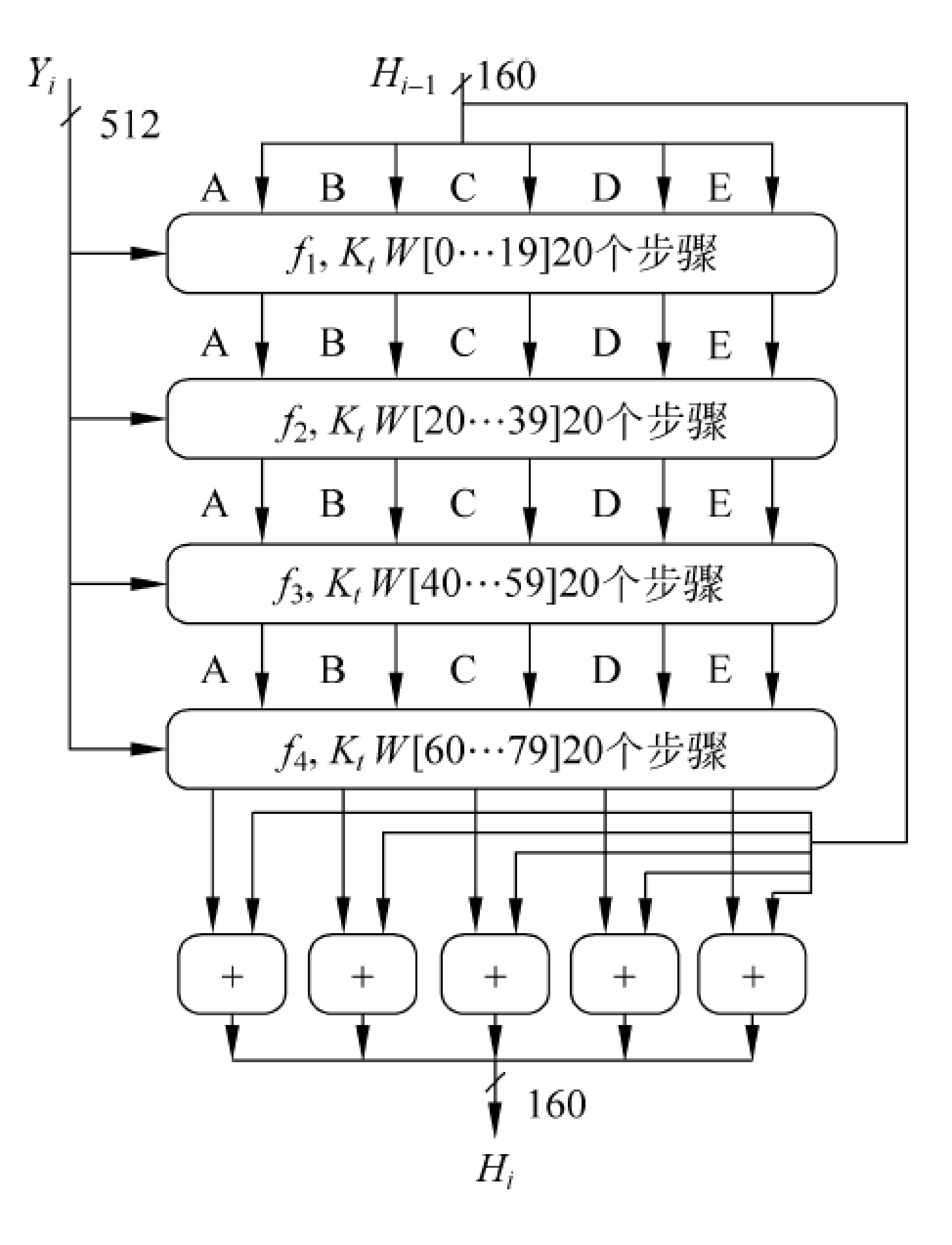


图 3-17　SHA-1 对单个 512 位分组的处理过程

图中的 f_1、f_2、f_3、f_4 为 4 个基本逻辑函数，它们的结构相似，每个循环使用不同的逻辑函

数。逻辑函数的定义为:

(1) $f_1(t,b,c,d) = (b \wedge c) \mid (\bar{b} \wedge d), 0 \leqslant t \leqslant 19$。

(2) $f_2(t,b,c,d) = b \wedge c \wedge d, 20 \leqslant t \leqslant 39$。

(3) $f_3(t,b,c,d) = (b \wedge c) \mid (b \wedge d) \mid (c \wedge d), 40 \leqslant t \leqslant 59$。

(4) $f_4(t,b,c,d) = b \wedge c \wedge i, 60 \leqslant t \leqslant 79$。

K_t 为常量字,可用 16 进制表示如下:

(1) $K_t =$ 0x5A827999,$0 \leqslant t \leqslant 19$。

(2) $K_t =$ 0x6ED9EBA1,$20 \leqslant t \leqslant 39$。

(3) $K_t =$ 0x8F1BBCDC,$40 \leqslant t \leqslant 59$。

(4) $K_t =$ 0xCA62C1D6,$60 \leqslant t \leqslant 79$。

图 3-17 中的+也是 mod 2^{32};Y_i 是指当前 512 位的消息分组;$W[j]$是由当前消息分组 Y_i 生成的一组字,总共 80 个,其生成规则为:$W[0]$~$W[15]$直接取自当前消息分组 Y_i 对应字的值,其他的定义如下:

$$W[t] = S^1(W[t-1] \oplus W[t-14] \oplus W[t-8] \oplus W[t-3])$$

其中 S^1 表示循环左移一位操作。

(5) SHA-1 的压缩函数如图 3-18 所示,A、B、C、D、E 这 5 个缓冲区的字分别用 a、b、c、d、e 表示,则 SHA-1 的压缩函数可表示为:

$$(a,b,c,d,e) \leftarrow ((e + f(t,b,c,d) + S^5(a) + W_t + K^t), a, S^{30}(b), c, d)$$

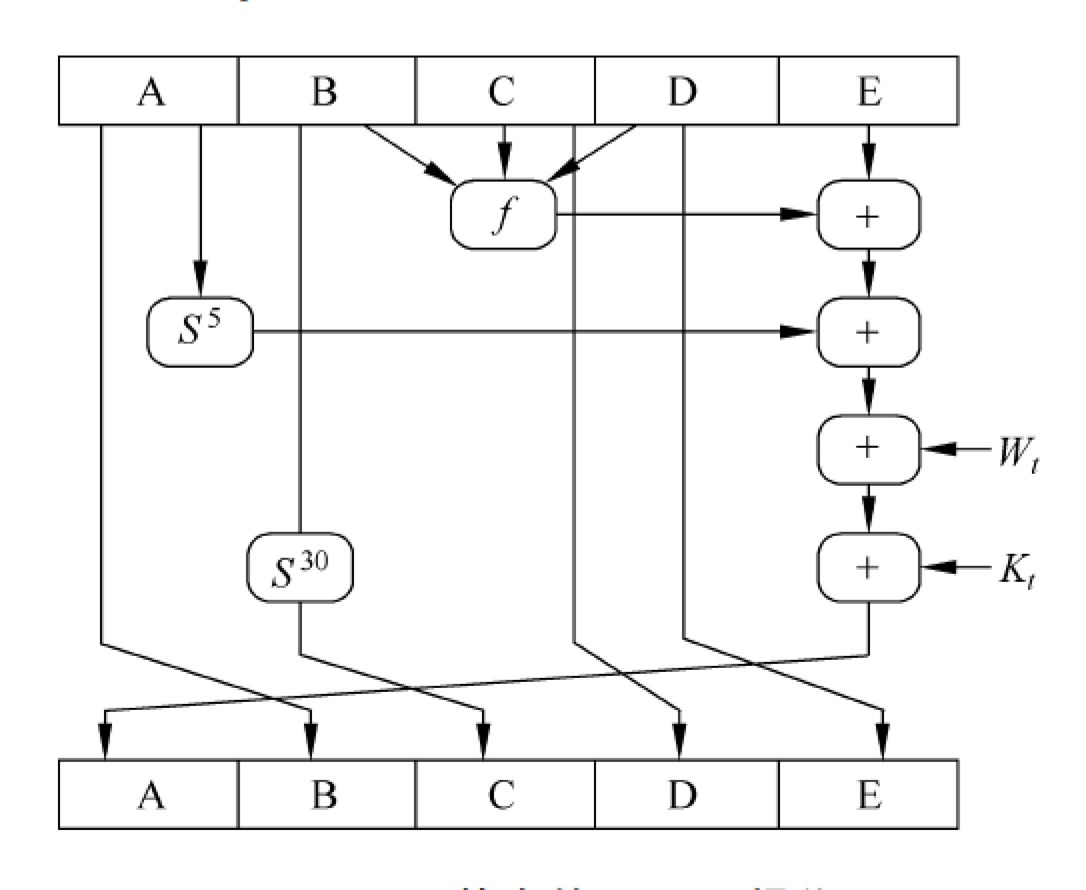


图 3-18 基本的 SHA-1 操作

3. Tiger hash

MD5 和 SHA-1 的结构比较简单,都由一些随机的转换组成。Tiger Hash 函数是由 Ross Anderson 和 Eli Biham 提出的,结构比 MD5 和 SHA-1 更复杂。事实上 Tiger 的结构更接近于分组密码。

同 MD5 和 SHA-1 一样,Tiger 的输入被分成 512 位的分组,若有需要则将输入填充为 512 位的整数倍。与 MD5 和 SHA-1 不同的是 Tiger 的输出是 192 位。选择输出位数为 192 的设计目的是适应 64 位处理器,因为 192 正好是 64 的倍数。在 Tiger 中每轮的中间值也是 192 位。

从它使用的 4 个 S 盒就可以看出 Tiger 的设计受到分组密码的影响,每个 S 盒将 8 位

映射成 64 位。Tiger 还应用了密钥扩展算法，这是因为没有密钥，实际上只是对输入分组进行扩展。

输入信息 X 被填充成 512 位的整数倍，然后写成

$$X = (X_0, X_1 \cdots, X_{n-1})$$

这里的每个 X_i 都是 512 位。Tiger 算法对每个 X_i 使用一个外循环，其中 $i=0$、1、2、…、$n-1$，每轮的结构如图 3-19 所示。

a、b、c 都是 64 位，第一轮初始值(a,b,c)是：

a=0x0123456789ABCDEF；

b=0xFEDCBA9876543210；

c=0xF096F5B4C3B2E187。

这里每轮的结果(a,b,c)作为下一轮的初始值。最后一轮的结果(a,b,c)就是 192 位 Hash 值。从这点上来看 Tiger 与分组密码非常相似。

注意：对于外循环 F_5 的输入是(a,b,c)。将 F_5 的输入标记为(a,b,c)，F_7 的输入标记为(c,a,b)。图 3-19 中每个函数 F_m 由 8 个如图 3-20 所示的内循环构成。将 512 位输入 W 写成：

$$W = (W_0, W_1, \cdots, W_7)$$

这里每个 W_i 都是 64 位。图 3-20 中每条线都代表 64 位。

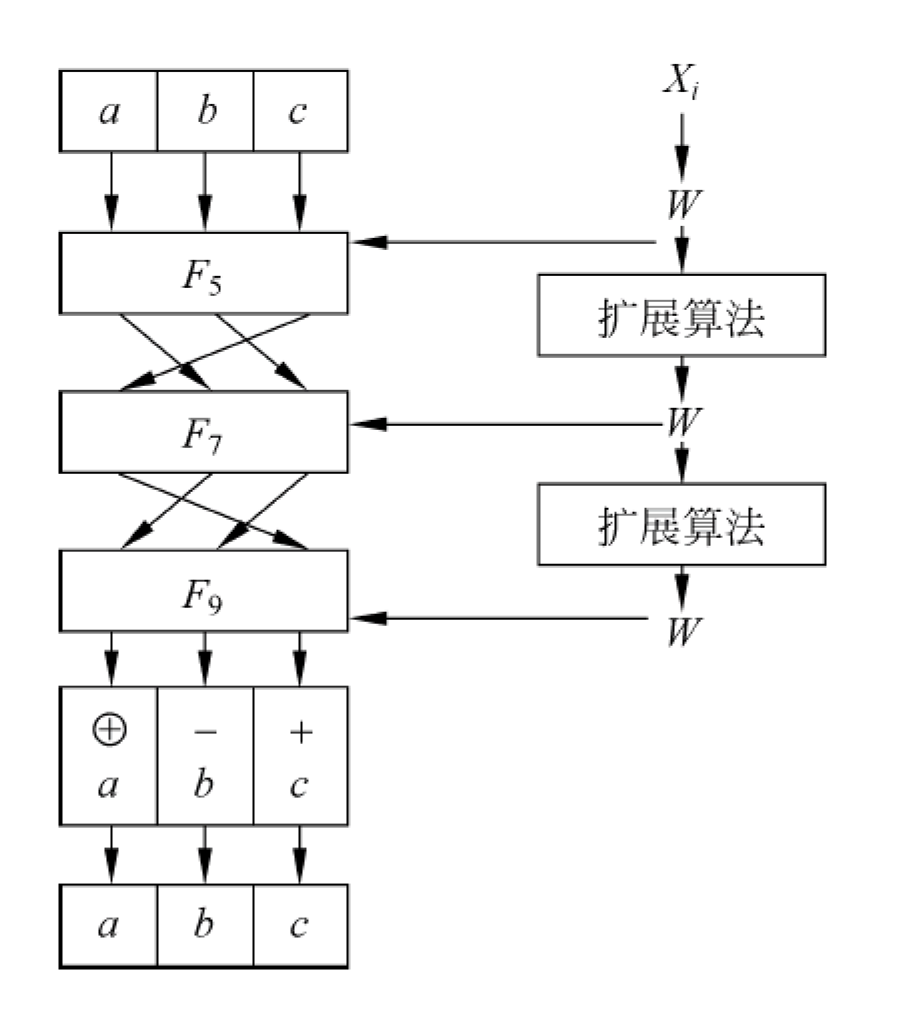



图 3-19　Tiger Hash 的外循环

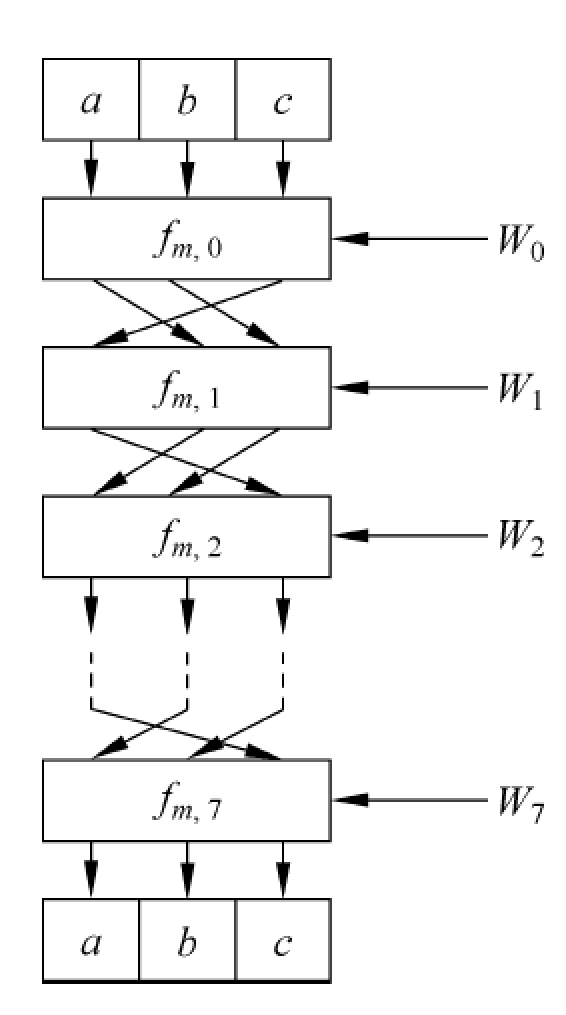


图 3-20　Tiger Hash 的 F_m 内循环

对于 $i=0$、1、2、…、7，$f_{m,i}$的输入值分别是：

$$(a,b,c),(b,c,a),(c,a,b),(a,b,c),(b,c,a),(c,a,b),(a,b,c),(b,c,a)$$

$f_{m,i-1}$的输出分别标记为(a,b,c)，每个 $f_{m,i}$依赖于 a、b、c、W_i 和 m，这里的 W_i 是 512 位输入 W 的第 i 个 64 位子块。$f_{m,i}$的下标 m 是乘数，定义如下：

将 c 写成

$$c = (c_0, c_1, \cdots, c_7)$$

这里每个 c_i 都是单字节。$f_{m,i}$定义如下：

$$c = c \oplus \omega_i$$

$$a = a - (S_0[c_0] \oplus S_1[c_2] \oplus S_2[c_4] \oplus S_3[c_6])$$
$$b = b - (S_3[c_1] \oplus S_2[c_3] \oplus S_1[c_5] \oplus S_0[c_7])$$
$$b = b \cdot m$$

这里每个 S_i 都是 8 位映射到 64 位的 S 盒。这些 S 盒很大,这里就不给出了。

4. CRC

CRC(Cyclic Redundancy Check,循环冗余校验码)由于实现简单,检错能力强,被广泛使用在各种数据校验应用中。CRC 占用系统资源少,用软硬件均能实现,是进行数据传输差错检测的一种很好的手段(CRC 并不是严格意义上的 Hash 算法,但它的作用与 Hash 算法大致相同,所以归于此类)。

生成 CRC 码的基本原理:任意一个由二进制位串组成的代码都可以和一个系数仅为 0 和 1 取值的多项式一一对应。例如,代码 1010111 对应的多项式为 $x^6+x^4+x^2+x+1$,而多项式 $x^5+x^3+x^2+x+1$ 对应的代码为 101111。

CRC 码集选择的原则:若设码字长度为 N,信息字段为 K 位,校验字段为 R 位($N=K+R$),则对于 CRC 码集中的任一码字,存在且仅存在一个 R 次多项式 $g(x)$,使得:

$$V(x) = A(x)g(x) = x^R m(x) + r(x) \tag{3-28}$$

其中,$m(x)$为 K 次原始的信息多项式;$r(x)$为 $R-1$ 次校验多项式(即 CRC 校验和,由多项式 $g(x)$对信息多项式 $m(x)$做模 2 除得到);$g(x)$称为生成多项式,其表达式为:

$$g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + g_{R-1}x^{R-1} + g_R x^R \tag{3-29}$$

发送方通过指定的 $g(x)$产生 CRC 码字,接收方则通过该 $g(x)$来验证收到的 CRC 码字。

CRC 校验码软件生成方法:借助于多项式除法,其余数为校验字段。

例 3-12 信息字段代码为 1011001,对应 $m(x)=x^6+x^4+x^3+1$。

假设生成多项式为:$g(x)=x^4+x^3+1$,则对应 $g(x)$的代码为 11001,则

$$x^4 m(x) = x^{10} + x^8 + x^7 + x^4$$

对应的代码记为:10110010000;

采用多项式除法,得余数为:1010 (即校验字段为 1010);

发送方发出的传输字段为:10110011010 (信息字段 校验字段);

接收方使用相同的生成码进行校验:接收到的字段与生成码进行二进制除法,如果能够除尽,则正确。下面给出余数(1010)的计算步骤:

除法没有数学上的含义,而是采用计算机的模二除法,即除数和被除数做异或运算。进行异或运算时除数和被除数最高位对齐,按位异或。

(1)
```
   10110010000
  ⊕11001
  ───────────
   01111010000
```

(2)
```
   1111010000
  ⊕11001
  ───────────
   0011110000
```

(3)　11110000

$\oplus$11001

00111000

(4)　111000

$\oplus$11001

001010

则 4 位 CRC 监督码就为：1010。

利用 CRC 进行检错的过程可简单描述为：在发送端根据要传输的 k 位二进制码序列，以一定的规则产生一个校验用的 r 位监督码(CRC 码)，附在原始信息后边，构成一个新的二进制码序列(共 $k+r$ 位)，然后发送出去。在接收端根据信息码和 CRC 码之间所遵循的规则进行检验，以确定传输中是否出错。这个规则在差错控制理论中称为生成多项式。

3.6.3　常见的消息认证码算法

MAC 实质上是一个将双方共享的密钥 k 和消息 m 作为输入的函数，如果将函数值记为 $\mathrm{MAC}_k(m)$，这个函数值就是一个认证标记，这里用 δ 表示。攻击者发起攻击的时候，所能得到的是消息和标记的序列对 $(m_1,\delta_1),(m_2,\delta_2),\cdots,(m_q,\delta_q)$(其中 $\delta_i=\mathrm{MAC}_k(m_i)$)。如果攻击者可以找到一个消息 m，m 不在 $m_1,\cdots,m_q$ 之中，并且能够得到正确的认证标记 $\delta=\mathrm{MAC}_k(m)$，就说明攻击成功了。攻击者成功的概率就是其攻破 MAC 的概率。

MAC 的构造方法有很多，主要有两种类型：一种是基于带密钥的 Hash 函数的；还有一种是基于流密码的，这种 MAC 不多也不流行，这里不作讨论。另外还有一种称为 Carter-Wegman MACs(首先使用一个泛 Hash 函数(universal Hash)将长消息散列成较短的字串，然后加密这个字串得到标记)，这种 MAC 基于的思想和前两类没有明确的界限，有些使用前两种方法构造的 MAC 也可看成是 Carter-Wegman MACs。这里主要讨论基于带密钥的 Hash 函数。

Hash 函数可以把较长的消息变换为较短的消息摘要，并且具有抵抗碰撞等良好的性质。为了保证消息的完整性，必须加入秘密信息——密钥，在加入了密钥之后，Hash 函数就称为带密钥的 Hash 函数。但是单独一个带密钥的 Hash 函数是不能直接作为 MAC 使用的，它必须经过特殊的构造，在具备了较好的性质后才可以用作消息认证码。在许多网络协议中，有很多使用这种方法构造的消息认证码，但是这些方法都是基于特殊技巧，很难进行安全性的分析和证明。这里介绍的基于带密钥的 Hash 函数的 MAC 可被证明是安全的。

基于带密钥的 Hash 函数的构造方法最早是由 M. Bellare 等人提出的，要求所使用的 Hash 函数具有迭代结构(例如 MD5、SHA-1 等)。所谓迭代结构就是反复地使用压缩函数 f 将长消息映射为短消息。这个压缩函数 f 具有两个输入，一个是长度为 l 的链变量，一个是长度为 b 的数据块，表示为 $f_k=f(k,x)$，其中 k 的长度为 l，x 的长度为 b。在 MD5 中 $b=512$，$l=128$。

假定消息 $x=(x_1,x_2,\cdots,x_n)$，其中 x_i 是长度为 b 的块，$i=1,2,\cdots,n$，n 是总块数。x_{n+1} 也是长度为 b 的二进制串，其中包含了 x 最后不足 b 的部分和整个消息长度的二进制表示以及一些填充位。使用压缩函数 f 构造的 Hash 函数 $F(x)$ 如图 3-21 所示。

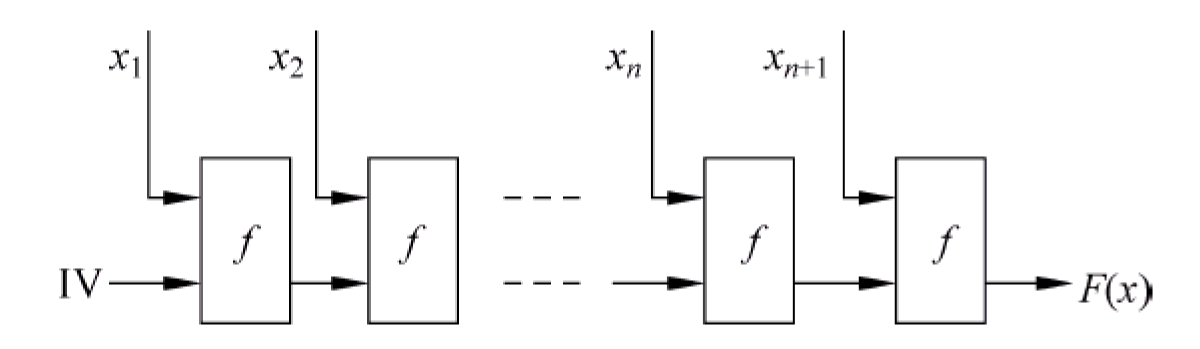


图 3-21 具有迭代结构的单向 Hash 函数

IV 代表初始向量,其长度为 l。如果使用密钥 k 作为初始向量,Hash 函数就成了带密钥的 Hash 函数。

如果让 F 代表初始向量固定为 IV 且具有迭代结构的 Hash 函数,那么 HMAC(x,k)的构造方法如表 3-14 所示。

其中,x 是任意长度的输入,k 是长为 l 的密钥,$\bar{k}$ 是密钥 k 填充到 b 位长之后的值。Opad 和 Ipad 是两个固定的长度为 b 的串。Opad 是 0x36 的重复直到 b 位长,Ipad 是 0x5c 的重复直到 b 位长。符号 ‖ 表示将两个二进制串串连起来。

表 3-14 HMAC 的构造

function HMAC(x,k)
$t \leftarrow 0, \bar{k} \leftarrow \text{pad}(x), x \leftarrow (\bar{k} \oplus I\text{pad}) \parallel x$
$t \leftarrow F(x), \delta \leftarrow F((\bar{k} \oplus O\text{pad}) \parallel t)$
return δ

HMAC 已经取代了 RFC 1828 成为 IPSec 协议中的认证算法。这种构造方法具备很多优点,和同类型的 MAC 算法相比,它给出了安全性证明,将 MAC 的安全性归结到所使用的 Hash 函数上。在软件实现上,它要比使用分组密码构造的 MAC 快,并且效率特别高。从它的构造上可以看出,它以一种非常简单的方式使用带密钥的 Hash 函数,同底层的 Hash 函数相比,性能并没有降低多少。另外两个值得称道的优点是免费和黑盒。免费是指使用 Hash 函数不受法律限制,可以免费使用。黑盒是指可以将底层的 Hash 函数看成一个模块,可根据需要方便地进行更换。

同时该算法也存在着不足:首先它的安全性依赖于底层的 Hash 函数,而所使用的 Hash 函数有些是没有安全性证明的,所以不能保证这种方法的安全性;其次由于压缩函数是串行的,该构造方法不支持并行。

UMAC 是由 Black 等人使用 Carter-Wegman 提出的思想(首先使用泛 Hash 函数的哈希消息 M,然后再使用密码学的方法加密所得的哈希值,即为想得到的标记)构造的一种算法。在这里将其划分为使用带密钥的 Hash 函数构造的 MAC,是因为该算法同样使用了带密钥的 Hash 函数。该算法首先使用 NH HASH 函数充分利用计算机的计算特点将源消息变换为原来消息长度的 $2/n$(其中 n 为子密钥的个数),然后再对所产生的消息使用 1 HMAC SHA 进行 Hash 变换。这种算法被认为是下一代的 MAC,其优点是速度很快,缺点是对于变化的长度需要进行特殊的处理。

3.6.4 分组加密与消息认证码

基于分组密码设计的这一类 MAC 主要有 CBC-MAC、EMAC(加密的 CBC-MAC)、XOR-MAC、PMAC、XECB-MAC 和 OCB 等。

1. CBC-MAC

CBC-MAC 其实就是对消息使用 CBC 模式进行加密,取密文的最后一块作为认证标

记。具体的构造方法如表 3-15 所示。

其中，x 为消息，k 为密钥，F 为某种分组密码算法，δ 就是所产生的标记。

表 3-15　CBC-MAC 的构造方式

function　CBC-MAC(x,k)
$y_0 \leftarrow 0^l$，pad(x)
partition x into $x_1,\cdots,x_n$
for $i \leftarrow 1$ to n
$y_i \leftarrow F_k(x_i \oplus y_{i-1})$
$\delta \leftarrow y_n$
return δ

这种方法出现得较早，是一种经典的构造方法，其构造方法简单，底层加密算法具备黑盒的性质，可以方便地进行替换。后来的很多 MAC 算法都是对它的改进。但是 CBC-MAC 仅适用于对相同长度的消息进行认证，在消息长度变化的情况下是不安全的。这些在本章的参考文献[5]中都已经给出了证明，另外它的构造方法决定了该算法不支持并行计算。

为了克服 CBC-MAC 的上述弱点，Bellare 给出了 3 种对 CBC-MAC 的改进方法，分别是：Input-length key separation，Length-prepending 和 Encrypt last block。其中最有效的方法是最后一种，也就是 EMAC。它是由 RIPE Project 在 1993 年提出的，接着被列入 ISO 标准中。它的具体构造是：

$$\text{EMAC}_{E_{k_1},E_{k_2}}(x) = E_{k_2}(\text{CBC}_{E_{k_1}}(x))$$

其中，k_1、k_2 是密钥空间中两个不同的密钥。通信双方使用一个安全的密钥 K 产生这两个密钥：

$$k_1 = E_k(0^l),\quad k_2 = E_k(1^l)$$

并且在证明这个 MAC 的安全性的时候认为 E_{k_1}、E_{k_2} 是两个独立随机选择函数。随后对 EMAC 进行改进，得到了 3 种新的 MAC 算法：ECBC、FCBC 和 XCBC。

2. XOR-MAC

XOR-MAC 有两种方式：无状态（XMACR）和有状态（XMACC）。这种算法在计算过程中引入索引值使得分组密码每次加密的明文各不相同，最后再将所有的密文异或。具体的构造方法描述如下：

假定 $|x|$ 代表消息 x 的长度（即包含多少位），并且它是 32 的倍数。

$$x = (x_1, x_2, \cdots, x_n)$$

其中，$|x_i| = 32, i = 1, 2, \cdots, n$。假定 $n < 2^{31}$。$<i>$ 是数字 i 的长度为 b 的二进制表示，代表块的索引号。发送者维持一个长度为 63 位的记数 r，在 XMACC 模式下它的初始值为 0，每次增加 1。在 XMACR 模式下，r 是随机选取的一个长度为 63 位的串。它们的具体构造方式如表 3-16 所示。

表 3-16　XMACR 和 XMACC 的构造

function XMACR(x,k)	function XMACC(x,k)
pad(x) $r \leftarrow \{0,1\}^{63}$	pad(x) ctr$\leftarrow$ctr$+1$
$y_0 = F_k(0 \parallel r)$	$y_0 = F_k(0 \parallel \text{ctr})$
partition x into $x_1,\cdots,x_n$	partition x into $x_1,\cdots,x_n$
for $i=1$ to n	for　$i=1$ to n
$y_i = y_{i-1} \oplus F_k(1 \parallel <i> \parallel x_i)$	$y_i = y_{i-1} \oplus F_k(1 \parallel <i> \parallel x_i)$
return　(r, y_n)	return　(ctr, y_n)

由于 XOR-MAC 使用异或来生成标记，这就为其带来了并行性、增量式、乱序验证（在验证的时候无须按照一定的顺序进行）等优点。在计算速度方面，基于 DES 的 XOR-MAC 在硬件实现效率上高于 CBC-MAC；在软件实现上使用 MD5 实现 XOR-MAC 效率较高。在安全性方面，它的安全性要高于 CBC-MAC。对攻击者来说，在理想情况下其攻击 XOR-MAC 成功的概率要比攻击 CBC-MAC 成功的概率低，并且这个概率跟消息长度没有关系。该算法的缺点是在算法中引入了索引信息，引起了消息的扩展，导致了加密次数的成倍增加，降低了运算速度。

3. PMAC

PMAC 可以看成是对 XOR-MAC 的改进，它也采用了异或的方法来得到 MAC，因此它具有可并行、支持消息的添加、截短和替换等优点。不过它与 XOR-MAC 有两点不同：第一是所加密的内容不同，XOR-MAC 所加密的内容是消息连接上一个索引信息，而 PMAC 使用的是消息和不同的串进行异或之后的值；第二是对最后一块消息的处理不同，XOR-MAC 并不对最后一块消息进行特殊处理，而 PMAC 并不直接加密最后一块，而是先填充然后和前面块的加密结果进行异或之后再分情况处理，最后再加密一次。该算法使用了灰码（Gray Code）和有限域 GF(2)上的乘法。

PMAC 的产生是在线的，也就是说在计算 MAC 的时候无须事先知道消息的长度。另外，PMAC 是确定性的，它不需要一个随机序列或维持一个计数。虽然 PMAC 具备这么多的优点，但是它的速度比 CBC-MAC 要慢，且该算法受专利保护，不能免费使用。

4. XECB-MAC

XECB-MAC 也可看成是 XOR-MAC 的一种改进，它仍然采用异或的方法得到 MAC，因此具有 XOR-MAC 的优点，例如支持并行计算、增量式操作、乱序验证等特性。和 XOR-MAC 不同的是它没有使用消息的有效位来记录消息的位置，这样就减少了加密的次数，因此它的速度要高于 XOR-MAC，但低于 CBC-MAC。而且它的安全性没有 XOR-MAC 的高。

由于在许多需要加密的情况下也同时需要对消息进行认证，而简单的将加密算法和认证算法结合起来的方法并不能保证其安全性。所以就出现了同时提供加密和认证的模式，这种模式有 XCBC 和 OCB 等。XCBC 对消息同时提供加密和认证，它也分为无状态和有状态两种。该方法支持实时的消息认证，所谓实时是指当加密完成时，认证标记就产生了。此外，该方法还具有支持并行计算等优点。该方法的不足之处在于使用了两个密钥，这给密钥的存储和分发带来了困难；而且所提供的完整性服务仅仅是对加密的一种补充，如果作为 MAC 单独使用，则会造成计算资源的浪费。

5. OCB

OCB 是在综合了 PMAC 和 XCBC-MAC 的构造方法的基础上提出来的，它同时提供了加密和认证。从它的构造方法上可以看出它与 PMAC 有一定的渊源，区别在于 OCB 同时提供加密和认证，而 PMAC 仅仅提供认证。OCB 的优点有很多，例如它能处理任意长度的消息、运算速度快并且支持并行处理。该模式在同时需要保证消息的私密性和完整性的情况下适用，例如可以用在 SSL 和 SSH 协议中以取代当前使用的组合算法。OCB 的缺点在于算法复杂并且受专利保护，不可免费使用。

3.7　数字签名技术

3.7.1　基本概念

数字签名(Digital Signature,又称公钥数字签名或电子签章)是一种类似写在纸上的普通的物理签名,它使用了公钥加密领域的技术实现,用于鉴别数字信息。一套数字签名通常定义为两种互补的运算,一个用于签名,另一个用于验证。数字签名了的文件的完整性是很容易验证的(无须骑缝章、骑缝签名,也无须笔迹专家),而且数字签名具有不可抵赖性。

简单地说,所谓数字签名就是附加在数据单元上的一些数据,或是对数据单元所做的密码变换。这种数据或变换允许数据单元的接收者用以确认数据单元的来源和数据单元的完整性并保护数据,防止被人(例如接收者)伪造。它是对电子形式的消息进行签名的一种方法,一个签名消息能在一个通信网络中传输。基于公钥密码体制和私钥密码体制都可以获得数字签名,目前主要是基于公钥密码体制的数字签名,包括普通数字签名和特殊数字签名。普通数字签名算法有 RSA、ElGamal、Fiat-Shamir、Guillou-Quisquarter、Schnorr、DES/DSA、Ong-Schnorr-Shamir 数字签名算法,另外还有椭圆曲线数字签名算法和有限自动机数字签名算法等。特殊数字签名有盲签名、代理签名、群签名、不可否认签名、公平盲签名、门限签名、具有消息恢复功能的签名等,它与具体应用环境密切相关。显然,数字签名的应用涉及法律问题,美国联邦政府基于有限域上的离散对数问题制定了自己的数字签名标准(DSS)。

数字签名技术是不对称加密算法的典型应用。数字签名的应用过程是:数据源发送方使用自己的私钥对数据校验和或其他与数据内容有关的变量进行加密处理,完成对数据的合法“签名”;数据接收方则利用对方的公钥来解读收到的“数字签名”,并将解读结果用于对数据完整性的检验,以确认签名的合法性。数字签名技术是在网络系统虚拟环境中确认身份的重要技术,完全可以代替现实过程中的亲笔签名,在技术和法律上有保证。在数字签名应用中,发送者的公钥可以很方便地得到,但他的私钥则需要严格保密。

3.7.2　常用的数字签名体制

关于 RSA 算法在非对称加密算法中已经具体介绍过,在此详细介绍 DSS 和 DSA 算法。

1. DSS

DSS 使用的是只提供数字签名的算法,与 RSA 不同的是,DSS 是一种公钥方法,但不能用于加密或密钥分配。图 3-22 对用 DSS 和 RSA 这两种数字签名的产生方法进行了对比。在 RSA 方法中,Hash 函数的输入是要签名的消息,输出是定长的 Hash 码,用发送方的私钥对该 Hash 码加密形成签名,然后发送消息及签名,接收方用发送方的公钥对签名进行解密,如果计算出的 Hash 码与解密出的结果相同,则认为签名是有效的。因为只有发送方拥有私钥,因此只有发送方能够产生有效的签名。

DSS 方法也是用 Hash 函数,它产生的 Hash 值和为此次签名而产生的随机序列 k 作为

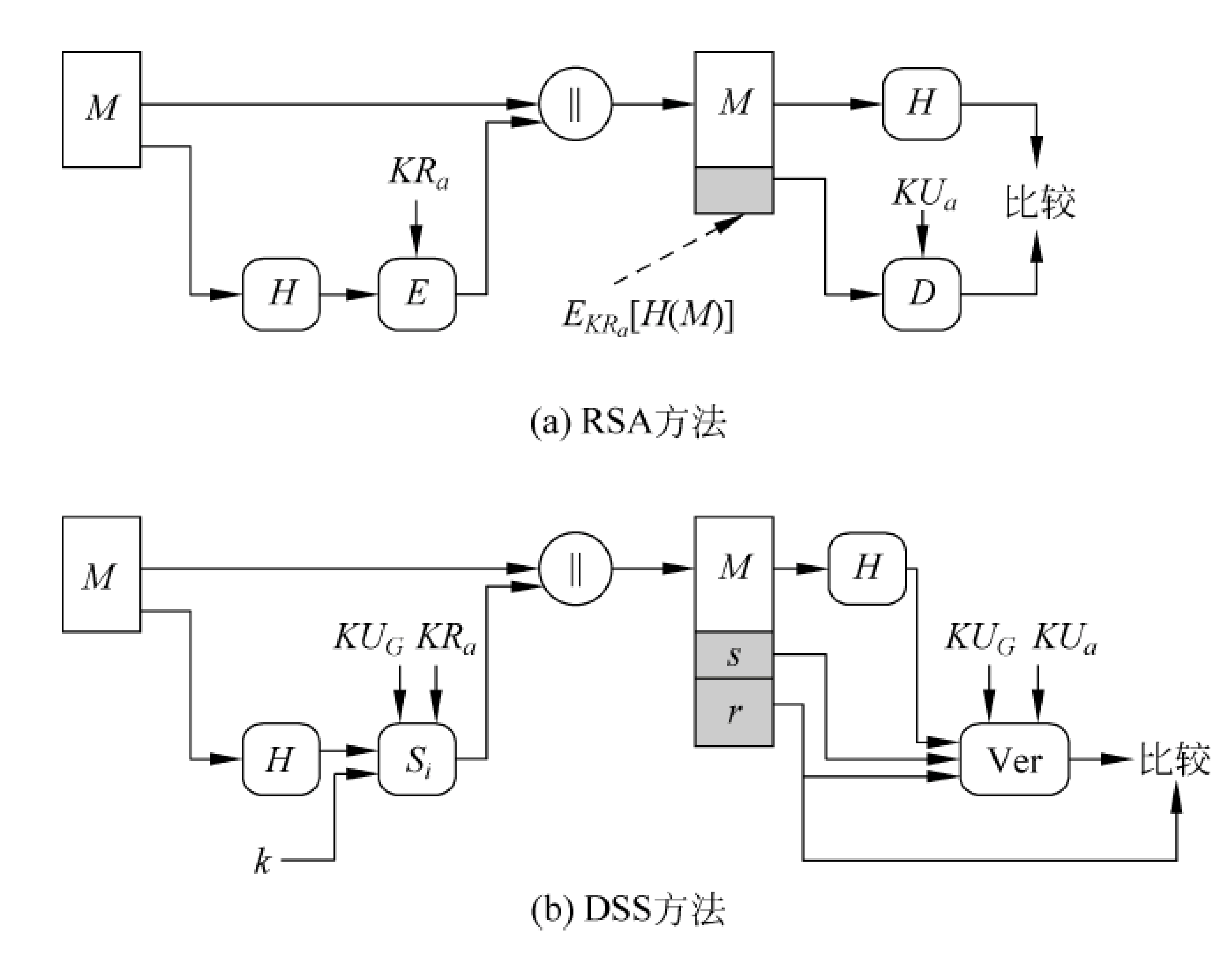


图 3-22 两种数字签名方法

签名函数的输入，签名函数依赖于发送方的私钥（KR_a）和一组参数，这些参数为通信多方所共有，可以认为这组参数构成的全局公钥（KU_a）签名由两部分组成，分别记为 s 和 r。

接收方对接收到的消息产生 Hash 码，这个 Hash 码和签名一起作为验证函数的输入，验证函数依赖于全局公钥和发送方公钥，若验证函数的输出等于签名中的 r 成分，则签名是有效的。签名函数保证只有拥有私钥的发送方才能产生有效签名。

2. DSA

DSA 建立在求离散对数的困难性以及 ElGamal 和 Schnorr 最初提出的方法之上。图 3-23 归纳总结了 DSA 算法，其中有 3 个公开参数为一组用户所共用。选择一个 160 位的素数 q，然后选择一个长度在 512～1024 且满足 q 能整除（$p-1$）的素数 p，最后选择 $h^{(p-1)/q} \bmod p$ 的 g，其中 h 是 1～（$p-1$）的整数，并且 g 大于 1。

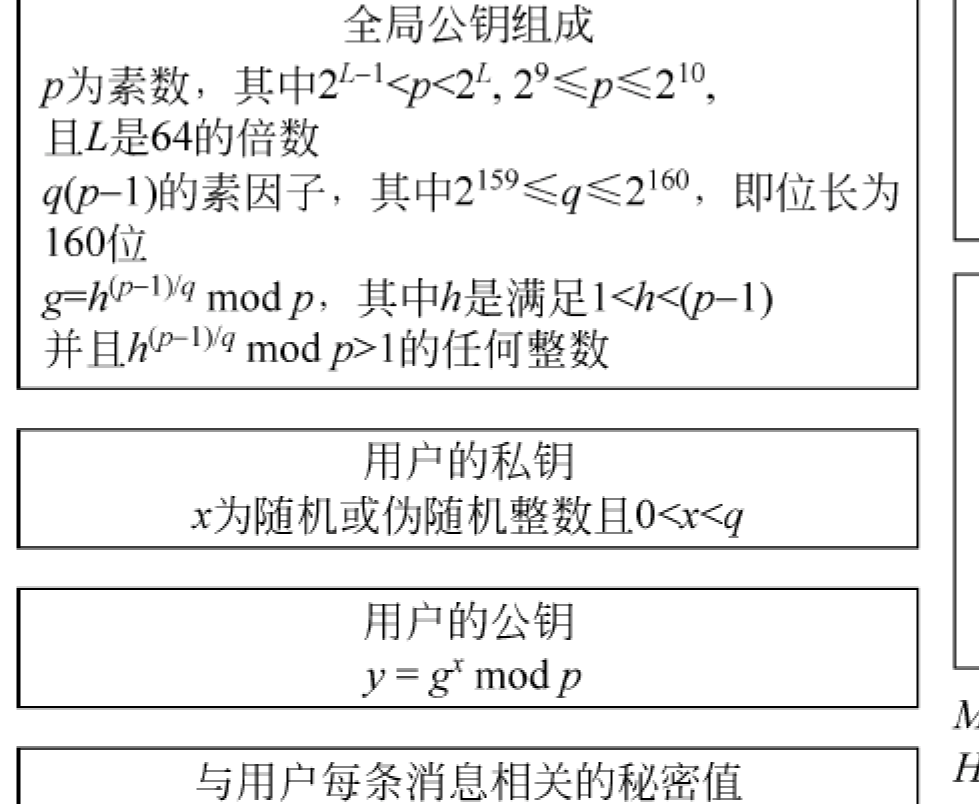


图 3-23 数字签名算法（DSA）

选定这些参数后，每个用户选择私钥并产生公钥。私钥 x 必须是随机或伪随机选择的在 1～$(q-1)$的数，可通过 $y=g^x \bmod p$ 计算得到公钥。由给定的 x 计算 y 比较简单，而由给定的 y 计算 x 则在计算上不可行，这就是求 y 的以 g 为底的模 p 的离散对数。

要进行签名，用户需计算两个量 r 和 s，r 和 s 是公钥(p,q,g)、用户私钥(x)、消息的 Hash 码 $H(M)$和附加整数 k 的函数，其中 k 是随机或伪随机产生的，且 k 对每次签名是唯一的。

图 3-24 更加详细地描述了上述签名和验证函数。该算法的特点为：接收端的验证依赖于 r，但是 r 根本不依赖于消息，它是 k 和全局公钥的函数。k 模 p 的乘法逆元传给函数 f_1，f_1 的输入还包含消息 Hash 值和用户私钥。函数的这种结构使接收方可利用其收到的消息和签名、它的公钥以及全局密钥来恢复 r。

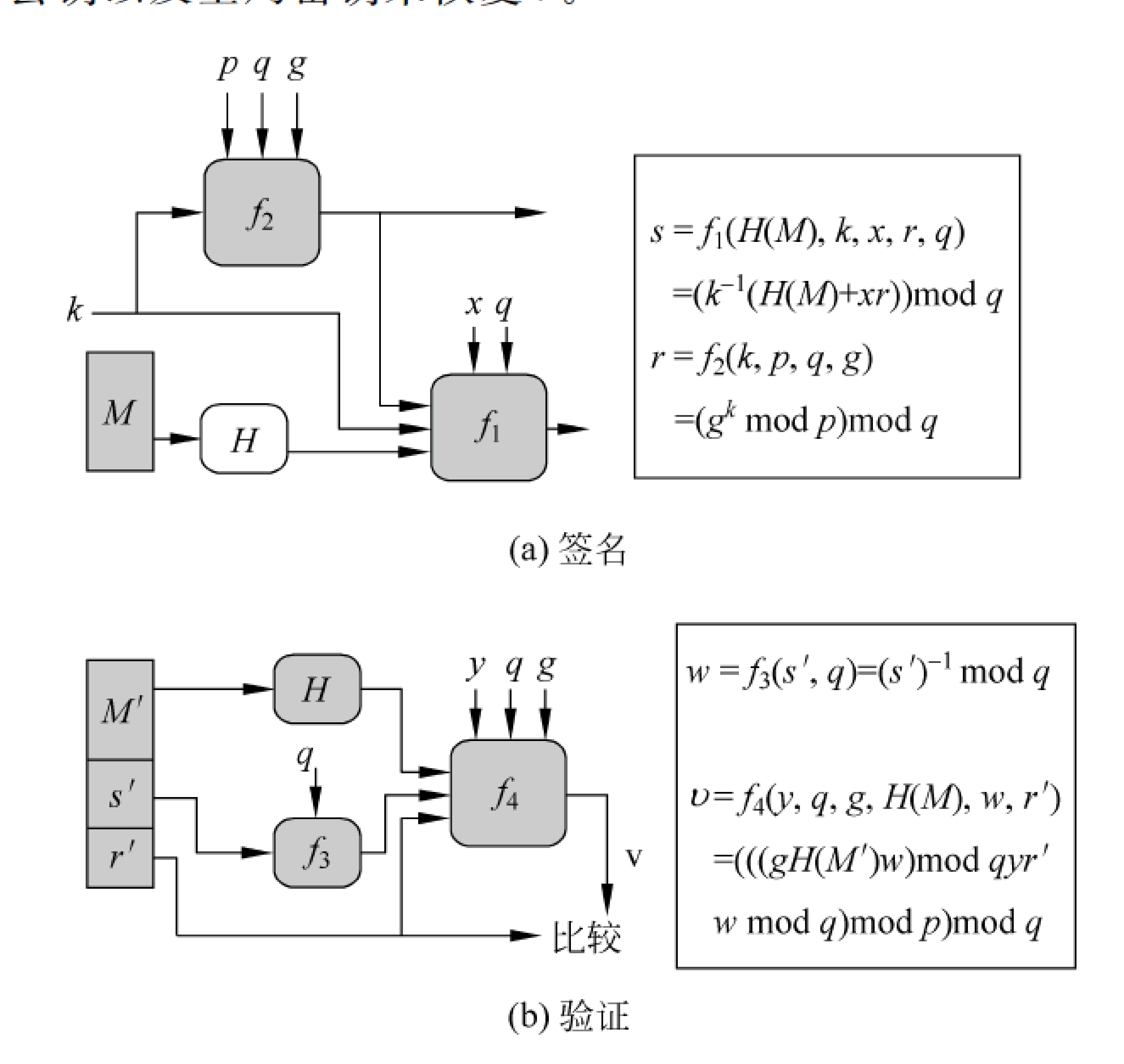


图 3-24　DSA 签名和验证

3.7.3　盲签名和群签名

1. 盲签名

一般的数字签名中，总是要先知道了文件内容后才签署，但有时需要对一个文件签名而不想让签名者知道文件的内容，称这样的签名为盲签名(Blind Signature)。盲签名最早是在 1982 年提出的。盲签名因为具有盲性这一特点，可以有效保护所签署消息的具体内容，所以在电子商务和电子选举等领域有着广泛的应用。

盲签名允许消息拥有者先将消息盲化，然后让签名者对盲化的消息进行签名，最后消息拥有者对签名除去盲因子，得到签名者关于原消息的签名。盲签名实际上就是接收者在不让签名者获取所签署消息具体内容的情况下所采取的一种特殊的数字签名技术，它除了满足一般的数字签名条件外，还必须满足以下两条性质：

(1) 签名者对其所签署的消息是不可见的，即签名者不知道他所签署消息的具体内容。

(2) 签名消息不可追踪，即当签名消息被公布后，签名者无法知道这是他哪次签署的。

对于盲签名一个非常直观的说明是：所谓盲签名，就是先将隐蔽的文件放进信封里，而除去盲因子的过程就是打开这个信封，当文件在一个信封中时任何人不能读它。对文件签名就是通过在信封里放一张复写纸，签名者在信封上签名时，他的签名便透过复写纸签到文件上。盲签名过程如图 3-25 所示。

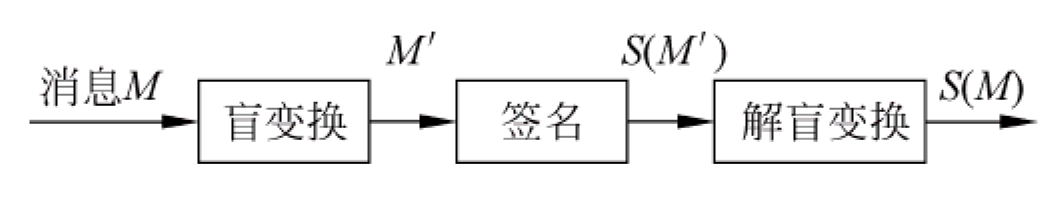


图 3-25 盲签名过程

现在假设 B 是担任仲裁人的角色，A 要求 B 签署一个文件，但是并不想让他知道文件的内容，只要求在需要时能够进行公正的仲裁。以下的协议就是实现这个签名的具体内容。

(1) 盲变换。A 将要签名的文件和一个随机序列(盲因子)相乘，这实际上完成了原文件的隐藏，隐藏后的文件称为盲文件。

(2) A 将该盲文件送给 B。

(3) B 对该文件签名。

(4) 解盲变换。A 对已签名的盲文件除以用到的盲因子，就得到 B 对原文件的签名。

只有当签名算法和乘法是可以交换的，上述的协议才可以真正实现，否则就要考虑用其他方法对原文件进行盲变换。为保证 B 不能进行欺诈活动，要求盲因子是真正的随机因子，这样 B 不能对任何人证明对原文件的签名，而只是知道对其签过名，并能验证该签名。这就是一个完全盲签名的过程。

一般来说，一个好的盲签名应该具有以下的性质：

(1) 不可伪造性。除了签名者本人外，任何人都不能以他的名义生成有效的盲签名，这是一条最基本的性质。

(2) 不可抵赖性。签名者一旦签署了某个消息，他无法否认自己对消息的签名。

(3) 盲性。签名者虽然对某个消息进行了签名，但他不可能得到消息的具体内容。

(4) 不可跟踪性。一旦消息的签名公开后，签名者不能确定自己何时签署的这条消息。

满足上面几条性质的盲签名，被认为是安全的。这 4 条性质既是设计盲签名所应遵循的标准，又是判断盲签名性能优劣的根据。

另外，方案的可操作性和实现的效率也是设计盲签名时必须考虑的重要因素。一个盲签名的可操作性和实现速度取决于以下几个方面：

(1) 密钥的长度。

(2) 盲签名的长度。

(3) 盲签名的算法和验证算法。

2. 群签名

群签名(Group Signature)是由 Chaum 和 van Heyst 在 1991 年提出的一个比较新的签名概念。Camenish、Stadler、Tsudik 等人对这个概念进行了修改和完善。群签名在管理军事、政治及经济等多个方面有着广泛应用。

所谓群签名就是满足这样要求的签名：在一个群签名方案中，一个群体中的任意一个成员可以以匿名的方式代表整个群体对消息进行签名。与其他数字签名一样，群签名是可

以公开验证的，而且可以只用单个群公钥来验证。也可以作为群标志来展示群的主要用途、种类等。群签名具有以下几个特点：

(1) 只有群体中的成员能代表群体签名。

(2) 接收到签名的人可以用公钥验证群签名，但是不可知道由群体中哪个成员所签。

(3) 发生争议时可由群体中的成员或者可信赖机构识别群中的签名者。

群签名有下面几个研究方向：

(1) 如何安全有效地废除群成员。即如何设计一个废除群成员的方法，使得一个群成员被删除后，原来的私钥和成员证书不能再用于签名，而且不影响他原来所做的签名的安全性。现有的群签名方案都不能安全有效地废除群成员。

(2) 如何设计高效地打开签名的算法。即如何使群管理员无须大的计算量就可以打开签名而确定出签名人的身份。

(3) 寻找一些安全高效的新的群签名算法。现有的相对安全高效的群签名方案基本上都依赖于RSA签名体制、Schnorr签名体制以及双重离散对数、离散对数的方根、有限循环群中元素的表示，某一秘密数值在一个指定的区间内的知识签名，效率都不是很理想。因此，寻求新的安全高效的群签名算法是很有必要的。

(4) 如何在电子商务等领域更广泛地使用群签名。在现有的文献中，关于群签名在电子商务领域的应用还不多见。由于群签名对于签名人能提供好的匿名性，同时又能使群管理员在必要的时候可以打开签名而撤销匿名性，所以可以广泛地应用于电子商务中的许多方面。只要找到高效使用的群签名算法，群签名在电子商务中的应用必然会走向实用。

(5) 对于群签名相关的数字签名及其应用的研究。目前对与群签名相关的数字签名及其应用的研究还不够，分级群签名、群盲签名、多群签名等都有实际应用背景，然而对它们的研究才处于起步阶段。

3.8　身份认证技术

前面介绍的数字签名和消息认证技术最主要的应用领域就是身份认证。本章首先阐述身份认证的基本概念，然后介绍典型的身份认证技术。

3.8.1　基本概念

身份认证是指定用户向系统出示自己身份的证明过程，通常是为了获得系统服务所必须通过的第一道关卡。在具有安全机制的系统中，任何一个想要访问系统资源的人都必须首先向系统证实自己的合法身份，然后才能得到相应的权限。在竞争激烈的现实社会中，为了获得非法利益，各种身份欺诈活动很频繁，因此在很多情况下都需要证明个人的身份。身份识别技术能使识别者识别到个人的真正身份，确保识别者的合法权益，这是社会责任的体现和社会管理的需要。

在网络环境下根据被认证方赖以证明身份的秘密的不同，身份认证可以基于如下因素：

(1) 根据个人所知道的信息来证明其身份：双方共享的数据，例如密码等。

(2) 根据个人所拥有的东西来证明其身份：被认证双方所拥有的外部物理实体，例如智能安全存储介质。

(3) 根据个人独一无二的身体特征来证明其身份：被认证双方所持有的生物特征，例如指纹、虹膜、语音、脸部等。

消息认证、数字签名和身份认证一同构成了整个鉴别系统。消息认证是对消息本身的鉴别，数字签名和身份认证是对发送消息的实体的鉴别。数字签名和身份认证都是确保消息发送方身份真实性的安全措施，数字签名也具有身份认证的功能，但身份认证同数字签名也有一些区别，主要表现在以下几个方面。

(1) 身份认证一般是基于发送方和接收方共享的秘密数据，以证实被鉴别对象的身份真实性。而用于验证签名的数据是公开的。

(2) 身份认证可以是单向认证或双向认证。所谓单向认证是指仅有一方需要对另一方进行身份鉴别；双向认证指的是通信双方互相进行身份鉴别。但数字签名则允许除发送方和接收方之外的第三者验证。

(3) 对于数字签名来说，发送方不能抵赖，接收方不能伪造，而身份认证却不一定具备这些特点。

(4) 身份认证技术的实现可能需要使用数字签名技术。

3.8.2 常用身份认证技术

1. 基于密码的身份认证

密码认证是根据用户的知识来进行身份的鉴别，该方法已经广泛应用于社会生活当中，这也是一种比较早的认证技术。当被认证对象要求访问提供服务的系统时，提供服务的认证方要求被认证对象提交密码信息，认证方收到对方的密码后，将其与系统中存储的用户密码进行比较，以确认被认证对象是否为合法访问者。这种认证方式叫做 PAP(Password Authentication Protocol，密码认证协议)。

密码生成主要有两种方式，即用户自己定义或由系统自动随机产生。前者的优点是用户比较容易记忆，但易被攻击者猜出；后者的优点是随机性好，不容易被猜出，但用户也不容易记忆。在实际的应用中，通常采用一些变形方式以防范对用户密码的猜测。

基于密码认证方法的优点在于，一般的系统如 UNIX、Windows NT、NetWare 等都提供了对密码认证的支持，应用对象广泛，特别对于封闭的小型系统来说不失为一种简单可行的方法。但是也正由于其简单性，基于密码认证方法存在着以下的不足：

(1) 用户以明文方式输入密码，很容易被内存中运行的黑客软件记录下来而泄密。

(2) 密码在传输过程中可能被截获。

(3) 窃取密码者可以使用字典穷举密码或者直接猜测密码。

(4) 攻击者可以利用服务系统中存在的漏洞获取用户密码。

(5) 密码的发放和修改过程都涉及很多安全性问题。

(6) 只能进行单向认证，即系统可以认证用户，而用户无法对系统进行认证。

针对 PAP 的问题，产生了挑战握手认证协议(Challenge Handshake Authentication Protocol，CHAP)，它采用挑战-应答(Challenge-Response)的方式，通过三次握手对被认证对象的身份进行周期性的认证。CHAP 加入不确定因素，通过不断地改变认证标识符和随

机的挑战消息来防止重放攻击，CHAP的认证过程如下：

(1) 被认证对象要求访问提供服务的系统时，认证方向被认证对象发送随机数据作为递增改变的标识符和一个挑战消息。

(2) 被认证对象向认证方发回一个响应，该响应数据由单向散列函数计算得出，单向散列函数的输入参数由本次认证的标识符、密钥和挑战消息构成。

(3) 认证方将收到的响应与自己根据认证标识符、密钥和挑战消息计算出的散列函数值进行比较。如果结果一致则认证通过，向被认证对象发送“成功”消息；否则发送“失败”消息并切断服务连接。

2. 双因子身份认证

双因子认证(Two-factor Authentication)是一种较为先进的身份认证技术，现在很多的身份认证系统都已经融入了双因子等先进技术。所谓双因子是指其中一个因子是只有用户本身知道的密码，它可以是个默记的个人认证号(Personal Identity Number，PIN)或密码，另一个因子是只有该用户拥有的外部物理实体——智能安全存储介质(智能卡)，将这两个因子相结合就成为双因子身份认证。

通常智能卡与一个密码或PIN结合使用，可存储用户个性化的秘密信息，同时验证服务器中也存放该秘密信息。认证时，用户输入个人身份识别信息(PIN)，智能卡认证PIN，若成功则读出卡中秘密信息，进而与主机之间认证。

与软盘、光盘等传统存储介质不同，智能安全存储介质具有Master Key和PIN密码及完善的信息加密和管理功能，非常适合作为身份认证应用保密信息的载体。它的主要优点是：

(1) 存储的信息无法复制。

(2) 具有双重密码保护机制和完备的文件系统管理功能。

(3) 某些智能安全存储介质还允许设置PIN猜测的最大值，以防止密码攻击。

基于双因子的认证方法比基于密码的认证方法增加了一个认证要素，因此该方法比基于密码的认证方法具有更好的安全性，在一定程度上解决了密码认证方法中的很多问题。

3. 生物特征认证技术

所谓生物特征认证技术是指通过计算机的强大功能，利用人体所固有的、唯一的、可靠稳定的生理特征或行为特征来进行个人身份认证或识别，因而具有更好的安全性、可靠性和有效性。基于生物特征的认证技术具有如下优点：不易遗忘或丢失；防伪性能好，一般很难伪造或被盗取；以人体特征做识别，“随身携带”，随时随地可用。

用于生物特征的认证技术主要有以下几种。

1) 指纹识别技术

指纹识别技术是最传统、最成熟的生物鉴定方式。由于没有两个人的皮肤纹路图样是完全相同的，即便是孪生儿也如此。指纹之所以能作为身份验证的准确而可靠的手段，是因为其具有独特的优势。

(1) 稳定性：从胎儿6个月时指纹完全形成到人死后，指纹的纹线类型、结构等始终不会有明显变化。

(2) 独特性：至今未找出两个指纹完全相同的人。根据指纹学理论，两枚指纹完全匹配上 12 个特征的几率为 10^{-50}。

(3) 便利性：提取指纹作为永久记录存档比较简单易行。

2) 声纹识别技术

声纹识别技术是指根据语音波形中反映说话人生理和行为特征的语音参数，自动识别说话人的身份。每个人说话的声音都会有自己的特点，人对语音的识别能力是特别强的。在商业和军事等安全性要求较高的系统中，常常靠人的声纹来实现个人身份的验证。

声纹识别与传统语音识别的区别：

(1) 声纹识别利用语音信号中的说话人信息，而无须考虑语音中的字词意思，它强调的是说话人的个性。

(2) 语音识别的目的是识别出语音信号中的言语内容，并不考虑说话人是谁，它强调共性。

3) 视网膜图样识别技术

人的视网膜血管的图样具有良好的个人特征，基于视网膜开发的识别系统在身份验证上有着独特的优势。视网膜识别的基本方法是用光学和电子仪器将视网膜血管图样记录下来，一个视网膜血管的图样可压缩为小于 35 字节的数字信息，可根据对图样的节点和分支的检测结果进行分类识别。

视网膜识别的验证效果相当好，但成本较高，运行的难度大(要求被识别人的合作并允许进行视网膜特征的采样)，因此只在军事或银行系统中被采用。

4) 虹膜图样识别技术

从理论上讲，虹膜认证是基于生物特征的认证中最好的一种认证方式。虹膜(眼睛中的彩色部分)是眼球中包围瞳孔的部分，上面布满极其复杂的锯齿网络状花纹，而每个人虹膜的花纹都是不同的。虹膜识别技术就是应用计算机对虹膜花纹特征进行量化数据分析，用以确认被识别者的真实身份。虹膜识别可以在 35～40 厘米的距离采样，比采集视网膜图样要方便，易为人所接受。基于虹膜的识别系统可用于安全入口、接入控制、信用卡、POS、ATM 等应用系统中，有效进行身份识别。

一个虹膜识别系统一般由 4 个部分组成：虹膜图像的采集、预处理、特征提取及模式匹配，如图 3-26 所示。

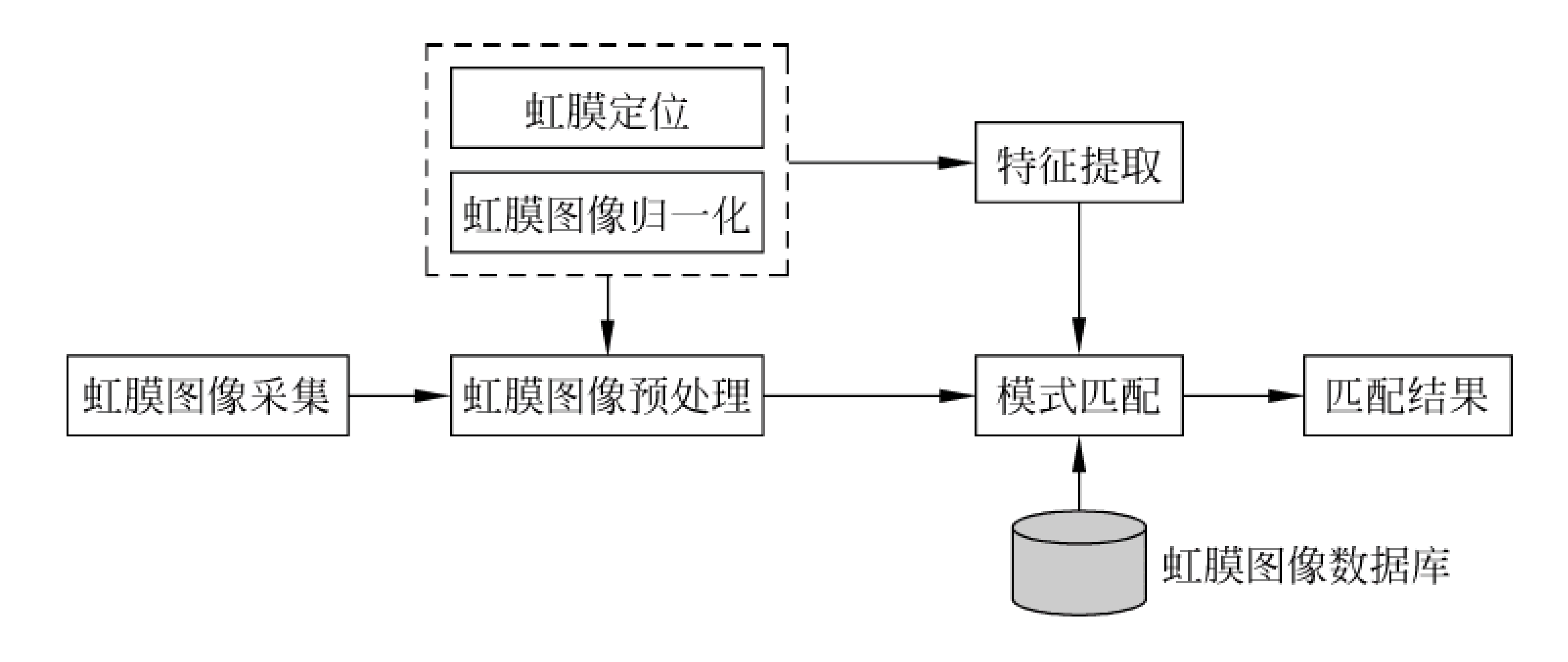


图 3-26　虹膜识别系统的组成

5）脸型识别技术

脸型识别技术是分析比较人脸视觉特征信息进行身份鉴别的技术。传统的脸部识别方法有两种：

（1）基于面部结构几何特征的脸像识别：这是最直观、最传统的方法，对人脸的描述非常紧凑，但存在特征提取困难、容易受头部姿势变化影响等缺点。

（2）基于模板匹配的方法：特征提取简单、准确度较高，但也很容易受到光照和姿势的影响。

思 考 题

（1）简述密码学的组成与分类。

（2）假设使用的密码是移动了 n 位的简单代替密码，试从下面的密文中找出明文和密钥：

AOPZPZHTLZZHNL

（3）用 Hill 加密消息：meet me at the usual place at then rather eight oclock，密钥为 $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$。要求写出计算过程和结果，并写出从密文恢复为明文所做的解密算法。

（4）当海军上尉 John F. Kennedy 下令日本毁灭者击沉美国巡逻号 PT-109 时，在澳大利亚的无线站截获了一条用 Playfair 密码加密的消息：

KXJEY UREBE ZWEHE WRYTU HEYFS
KREHE GOYFI WTTTU OLKSY CAJPO
BOTEI ZONTX BYBWT GONEY CUZWR
GDSON SXBOU YWRHE BAAHY USEDQ

密钥为 royal new Zealand navy。请解密这条消息。

（5）本题探讨 Vigenère 密码的一次一密版本的用途。在这种方案中，密钥是 0～26 的随机序列。例如，如果密钥是 3 19 5…，则密文的首个字母使用 3 个字母的移位加密，第二个字母使用 19 个字母的移位加密，第三个字母使用 5 个字母的移位加密，以此类推。

① 使用密钥流 9 0 1 7 23 15 21 14 11 11 2 8 9 加密明文 sendmoremoney。

② 使用①中产生的密文找到一个密钥，以便该密文解密为 cashnotneeded。

（6）实现 RC4 算法。假设密钥由下列 7 个字节构成：

key = (0x1A，0x2B，0x3C，0x4D，0x5E，0x6F，0x77)

① 列出初始化阶段之后的 S。

② 列出生成 100 字节的密钥流之后的置换 S。

③ 列出生成 1000 字节的密钥流之后的置换 S。

（7）Alice 的 RSA 公钥是$(N,e)=(33,3)$，私钥是 $d=7$。

① 如果 Bob 加密消息 $M=19$ 给 Alice，密文 C 是什么？展示 Alice 将 C 解密到 M 的过程。

② 令 S 是 Alice 对于消息 $M=25$ 的数字签名结果。S 是多少？如果 Bob 收到 M 和 S，

解释 Bob 验证签名的过程(假设这一签名过程验证成功)。

(8) 对于椭圆曲线:

$$E: y^2 = x^3 - 4(\bmod 211)$$

和曲线 E 上的点 $G=(2,2)$,假设使用 E 和 P 进行 ECC 的 Diffie-Hellman 密钥交换,这里 Alice 选择秘密值 $A=121$,Bob 选择秘密值 $B=203$,Alice 发送给 Bob 的值是什么? Bob 发送给 Alice 的值是什么? 共享的秘密是什么?

(9) 如何使用高级加密标准(AES)作为安全 Hash 函数? 注意 Hash 函数不适用密钥。

(10) 假设要加密一个由 3 块分组明文 P_0、P_1 和 P_2 组成的消息,只使用 Hash 函数和一个对称密钥 K,怎样对这个消息进行加密和解密?

(11) 找出下列两条消息(表示为十六进制)的所有不同的位,并验证它们的 MD5 Hash 值是相同的:

D1 31 DD 02 C5 E6 EE C4	69 3D 9A 06 98 AF F9 5C
2F CA B5 87 12 46 7E AB	40 04 58 3E B8 FB 7F 89
55 AD 34 06 09 F4 B3 02	83 E4 88 83 25 71 41 5A
08 51 25 E8 F7 CD C9 9F	D9 1D BD F2 80 37 3C 5B
96 0B 1D D1 DC 41 7B 9C	E4 D8 97 F4 5A 65 55 D5
35 73 9A C7 F0 EB FD 0C	30 29 F1 66 D1 09 B1 8F
75 27 7F 79 30 D5 5C EB	22 E8 AD BA 79 CC 15 5C
ED 74 CB DD 5F C5 D3 6D	B1 9B 0A D8 35 CC A7 E3

和

D1 31 DD 02 C5 E6 EE C4	69 3D 9A 06 98 AF F9 5C
2F CA B5 07 12 46 7E AB	40 04 58 3E B8 FB 7F 89
55 AD 34 06 09 F4 B3 02	83 E4 88 83 25 F1 41 5A
08 51 25 E8 F7 CD C9 9F	D9 1D BD 72 80 37 3C 5B
96 0B 1D D1 DC 41 7B 9C	E4 D8 97 F4 5A 65 55 D5
35 73 9A 47 F0 EB FD 0C	30 29 F1 66 D1 09 B1 8F
75 27 7F 79 30 D5 5C EB	22 E8 AD BA 79 4C 15 5C
ED 74 CB DD 5F C5 D3 6D	B1 9B 0A 58 35 CC A7 E3

(12) 因为 DSS 对每个签名产生一个 k,所以即使对同一消息在不同情况下的签名也不同,但 RSA 签名则不能做到这一点。这种区别有什么实际意义?

(13) 可以利用 Hash 函数构造类似 DES 结构的分组密码。但 Hash 函数是单向的,而分组密码是可逆的(解密),那么如何用 Hash 码构造上述的分组密码呢?

(14) 典型的身份认证技术有哪些? 简述其认证原理。

附:本章思考题答案或提示

(2) 使用穷举法,将密文每个字母向后移动 1、2、…、25 个位置,明文为 THISISAMESSAGE,密钥为 7。

(3) 将明文分组(m,e)、(e,t)、(m,e)、(a,t)、(t,h)、(e,u)、(s,u)、(a,l)、(p,l)、(a,c)、(e,a)、(t,t)、(h,e)、(n,r)、(a,t)、(h,e)、(r,e)、(i,g)、(h,t)、(o,c)、(l,o)、(c,k),再分别转

化为在字母表中的顺序，例如(m,e)转化为(13,5)。

$$C(m,e) = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} * \begin{bmatrix} 13 \\ 5 \end{bmatrix} (\bmod 26) = \begin{bmatrix} 23 \\ 15 \end{bmatrix} = (w,o)$$

解密即逆运算，模 26 逆矩阵：$A^{-1}(\bmod m) = |A|^{-1}(\bmod m) * A^{*}(\bmod m)$

模 26 倒数表(整数 a 有模 m 倒数的充要条件为 a 与 m 无公共素因子)如表 3-17 所示。

表 3-17　模 26 倒数表

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}(mod 26)	1	9	21	15	3	19	7	23	11	5	17	25

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}^{-1} (\bmod 26) = 3^{-1}(\bmod 26) * \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix} (\bmod 26) = \begin{bmatrix} 1 & 8 \\ 0 & 9 \end{bmatrix}$$

$$D(w,o) = \begin{bmatrix} 1 & 8 \\ 0 & 9 \end{bmatrix} * \begin{bmatrix} 23 \\ 15 \end{bmatrix} (\bmod 26) = \begin{bmatrix} 13 \\ 5 \end{bmatrix}$$

完成对明文分组(m,e)的加/解密。

(4) 按规则构造密钥如表 3-18 所示。

① 构造 5×5 矩阵(如表 3-18 所示)

表 3-18　思考题(4)的密钥表

R	O	Y	A	L
N	E	W	Z	D
V	B	C	F	G
H	I/J	K	M	P
R	S	T	U	X

② 将明文每两个一组，分组进行解密得：

PT BO AT ON EO WE NI NE LO ST IN AC TI ON IN BL AC KE TT ST RA IT TW OM IL ES SW ME RE SU CO VE XC RE WO FT WE LV EX RE QU ES TA NY IN FO RM AT IO NX

③ 将解密后的分组重新组合得：

PT BOAT ONE OWE NINE LOST IN ACTION IN BLACKETT STRAIT TWO MILES SW MERESU COVE X CREW OF TWELVE X REQUEST ANY INFORMATION X

(5) ① 按 Vigenère 密码加密规则，明文 sendmoremoney 加密为：BEOKJDMSXZPMH。

② 密钥为：25,4,22,3,22,15,19,5,19,21,12,8,4。

(6) ①可参考表 3-4 伪代码编程实现。

②、③可参考表 3-5 伪代码编程实现。

(7) ① $C = M^e \bmod n = 19^3 \bmod 33 = 28$，$M = C^d \bmod n = 28^7 \bmod 33 = 19$。

② $S = M^d \bmod n = 25^7 \bmod 33 = 31$

Bob 用 Alice 的公钥对收到的 M 进行签名得到 S'，将 S' 与收到的 S 相比较，相同则说

明 M 的内容没有经过篡改,验证成功,否则不成功。

(8) Alice 的公钥:A=121(2,2)=(115,48)

Bob 的公钥:B=203(2,2)=(130,203)

共享的密钥:121(130,203)=203(115,48)=(161,169)

(9) 提示:借鉴 Tiger 算法的外循环。

(10) 提示:采用基本的 Hash 函数,在发送端和接收端共享加密密钥 K,明文和 Hash 值被加密保护,提供保密和鉴别的功能。

(11) 提示:有 6 个字节不同,MD5 Hash 值均为 A4C0D35C95A63A805915367DCFE6-B751。

(12) 提示:RSA 既可以作为加密算法使用,又可以用作数字签名和密钥的分配与管理,而 DSA 只适合数字签名。

(13) 提示:将 Hash 函数构造方法里的压缩函数使用分组密码代替。

参考文献

[1] 陈鲁生,沈世镒. 现代密码学. 北京:科学出版社,2002.
[2] Douglas Stinson. Cryptography:Theory and Practice. Boca Raton:CRC Press,1995.
[3] D. E. R. Denning. Cryptography and Data Security. London:Addison-Wesley Puplishing Company,1982.
[4] Arto Salomaa. Public-key Cryptography. Springer,1990.
[5] Douglas R Slinson. 密码学原理与实践. 北京:电子工业出版社,2003.
[6] 冯登国,裴定一. 密码学导引. 北京:科学出版社,1999.
[7] P. Gardner. Marking,Breaking Codes :An Introduction to Cryptology. Prentice Hall,2001.
[8] S. Katzenbeisser. Information Hiding Techniques for Steganography and Digital Watermarking. Boston:Artech House,2000.
[9] S. Singh. The Code Book:The Science of Secrecy from Ancient Egypt to Quantum Cryptography. New York:Anchor Books,1999.
[10] R. K. Nichols. ICSA Guide to Cryptography. New York:McGraw-Hill,1999.
[11] A. Fernandes. Elliptic Curev Cryptography. Dr. Dobb's Journal,2001:56~63.
[12] 彭飞. 混沌密码算法及其在安全电子邮件中的应用. 华南理工大学博士学位论文,2006.
[13] 彭飞,龙敏,刘玉玲. 数字内容安全原理与应用. 北京:清华大学出版社,2012.
[14] 牛少彰,崔宝江,李剑. 信息安全概论. 北京:北京邮电大学出版社,2004.
[15] 郝玉洁,刘桂松,秦科. 信息安全概论. 成都:电子科技大学出版社,2007.
[16] 金澈明,汤云剑,等. 网络与信息安全技术. 上海:华东理工大学出版社,2009.
[17] 李伟超. 计算机信息安全技术. 长沙:国防科技大学出版社,2010.
[18] R. Jueneman,S. Matyas,C. Meyer. Message Authentication. IEEE Communications Magazine,1988.
[19] R. Juneman. Electronic Document Authentication. IEEE Network,1987,(1):17~23.
[20] A. Menezes, P. Oorshcot,S. Vanstone. Handbook of Applied Cryptography. Boca Raton:CRC Press,1997.
[21] D. Stinson. Cryptography:Theory and Practice. Boca Raton:CRC Press,2002.
[22] J. Black,P. Rogaway. A Block-Cipher Mode of Operation for Parallelizable Message Authentication. Advances in Cryptology-EUROCRYPT 2002,Heidelberg:Springer-Verlag,2002,384~401.
[23] D. G. Virgil, D. Pompiliu. Fast Encryption and Authentication:XCBC Encryption and XECB Authentication Modes. In FSE 2001 Yokohama Heidelberg:Springer-Verlag,2002,92~141.

[24]　P. Rogaway，M. Bellare，J. Black. OCB：A Block-Cipher Mode of Operation for Efficient Authenticated Encryption. ACM Conference on Computer and Communications Security（CCS '01），Philadelphia，PA，USA，2001，ACMPress，2001，196～205.

[25]　J. Deepakumara，H. M. Heys，R. Venkatesan. Performance comparison of message authentication code（MAC）algorithms for Internet protocol security（IPSEC）. Proc. Newfoundland Electrical and Computer Engineering Conf. ，St. John's，Newfoundland，Nov. 2003.

[26]　SG Akl. Digital Signatures：A Tutorial Survey. Computer，1983，(16)：15～24.

[27]　王大东，林东岱，吴文玲. 消息认证码研究. 通信和计算机，2005，(2)：76～81.

[28]　王海艳，王汝佳. 群签名方案之比较研究. 计算机应用研究，2005，(22)：93～95.

第4章　访问控制与VPN技术

本章学习目标

计算机网络访问控制技术是经过授权的主体向某些客体提供所授权的访问服务，同时拒绝向非授权的主体访问客体的行为提供服务的策略。它可以限制访问主体对关键资源的访问，防止非法用户的侵入或因合法用户的不慎操作而造成的破坏。本章对访问控制的概念、模型及访问控制中涉及的AAA组件进行介绍，然后详细介绍访问控制技术中的应用——虚拟专用网(Virtual Private Network，VPN)技术，VPN技术的隧道协议是本章的重点。

通过对本章的学习，应掌握以下内容：

(1) 访问控制技术的概念和特点。

(2) 访问控制分类。

(3) AAA技术。

(4) VPN的定义及其类型。

(5) VPN的各种隧道协议。

随着网络技术日益广泛而深入地应用到社会各个领域中的同时，网络安全问题却成为困扰和阻挠网络技术进一步普及、应用的绊脚石，如果不能有效地解决，必然会影响整个网络的发展。为此，国际标准化组织ISO在网络安全体系的设计标准ISO 7498—2中提出了层次型的安全体系结构，并定义了五大安全服务功能：身份认证服务，访问控制服务，数据保密服务，数据完整性服务，不可否认服务。作为五大服务之一的访问控制服务在网络安全体系结构中具有不可替代的作用，它可以限制对关键资源的访问，防止非法用户的侵入或合法用户的不慎操作所造成的破坏。

访问控制方法的研究一直是信息安全课题中的一个热门话题。本章主要介绍访问控制的概念、模型，重点介绍访问技术的分类、访问控制技术AAA和VPN技术及其隧道协议。

4.1　访问控制技术

4.1.1　访问控制技术的基本概念

访问控制(Access Control，AC)是针对越权使用资源的防御措施，即判断使用者是否有权限使用或更改某一项资源，并且防止非授权的使用者滥用资源。

访问控制通常包含以下三方面含义：一是机密性控制，保证数据资源不被非法读出；二是完整性控制，保证数据资源不被非法增加、改写、删除或生成；三是有效性控制，保证资源不被非法访问主体使用和破坏。访问控制对系统而言，可以有效地将未经授权的非法访问主体拒之于系统之外；对系统资源而言，则是保证资源不被非法访问和使用。以身份认证作为访问控制的前提，将各种安全策略相互配合才能真正起到信息资源的保护作用。

访问控制系统一般包括：

（1）主体（Subject）：发出访问操作、存取要求的主动方，通常指用户或用户的某个进程。

（2）客体（Object）：被调用的程序或欲存取的数据访问。

（3）安全访问政策：是一套规则，用以确定一个主体是否对客体拥有访问能力。

4.1.2　访问控制模型

在 GB/T18793.3（等同于 ISO/IEC10181-3）中定义了访问控制系统设计时所需的一些基本功能组件，并且描述了各功能组件之间的通信状态。

访问控制功能组件包括下列 4 个部分：发起者（Initiator）、访问控制实施功能（Access Control Enforcement Function，AEF）、访问控制决策功能（Access Control Decision Function，ADF）以及目标（Target），如图 4-1 所示。

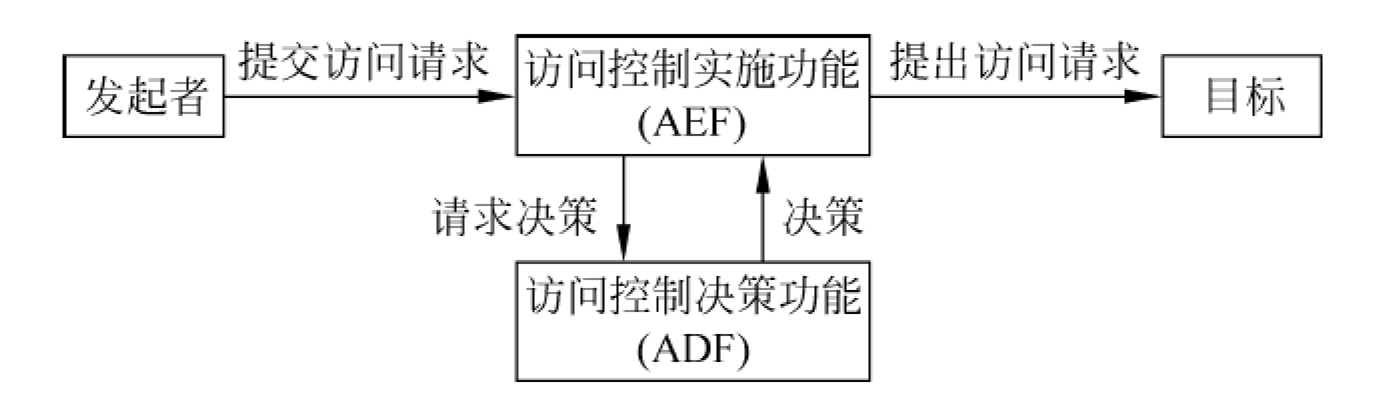


图 4-1　访问控制模型的基本组成

发起者是指信息系统中系统资源的使用者或是计算机程序等属于系统实体中主动的部分；目标是指被发起者所访问或试图访问的实体；AEF 的功能是负责建立发起者与目标之间的通信桥梁，它必须依照 ADF 的授权查询指示来实施上诉动作，也就是说，当发起者对目标提出执行操作要求（Access Request）时，AEF 会将这个请求信息通知 ADF，并由 ADF 做出授权准许的判决；ADF 根据 AEF 传输来的操作要求，以及访问控制决策信息（Assess Control Decision Information，ADI），作为判断访问控制的决策工作。

在信息系统中，ADF 是访问控制的核心。当 ADF 对发起者所传输的判断请求进行查核验证时，是根据不同来源端所送入的 ADI 以及其他附属作为判断的依据。这些不同来源的 ADI 包括以下内容。

（1）发起者 ADI：描述了发起者被赋予的执行权限。

（2）目标 ADI：描述了目标可被操纵的权限范围。

（3）访问请求 ADI（Access Request ADI）：是附带于访问控制请求时的决策信息。

（4）访问控制策略规则（Access Control Policy Rules）：是根据 ADF 所属的安全域机构，将所规范的政策转换为具体可被信息系统执行的规则。

（5）上下文信息（Contextual Information）：包括发起者的位置、访问时间或使用的特殊通信路径。

（6）保留的 ADI：是上次获准执行的相关信息，用于协助系统完成当前的核查任务。

4.1.3　访问控制组件的分布

在各种系统环境下，可以依照本身所制定的信息安全策略决定 AEF 的部署。例如，在

分布式系统中,AEF 所提供的功能可以独立成一个功能组件,称为 AEF Component,简称为 AEC。同样地,ADF 所提供的功能也可以独立成一个功能组件,称为 ADF Component,简称为 ADC。AEC 和 ADC 可以有不同的组合方式,如图 4-2 所示。

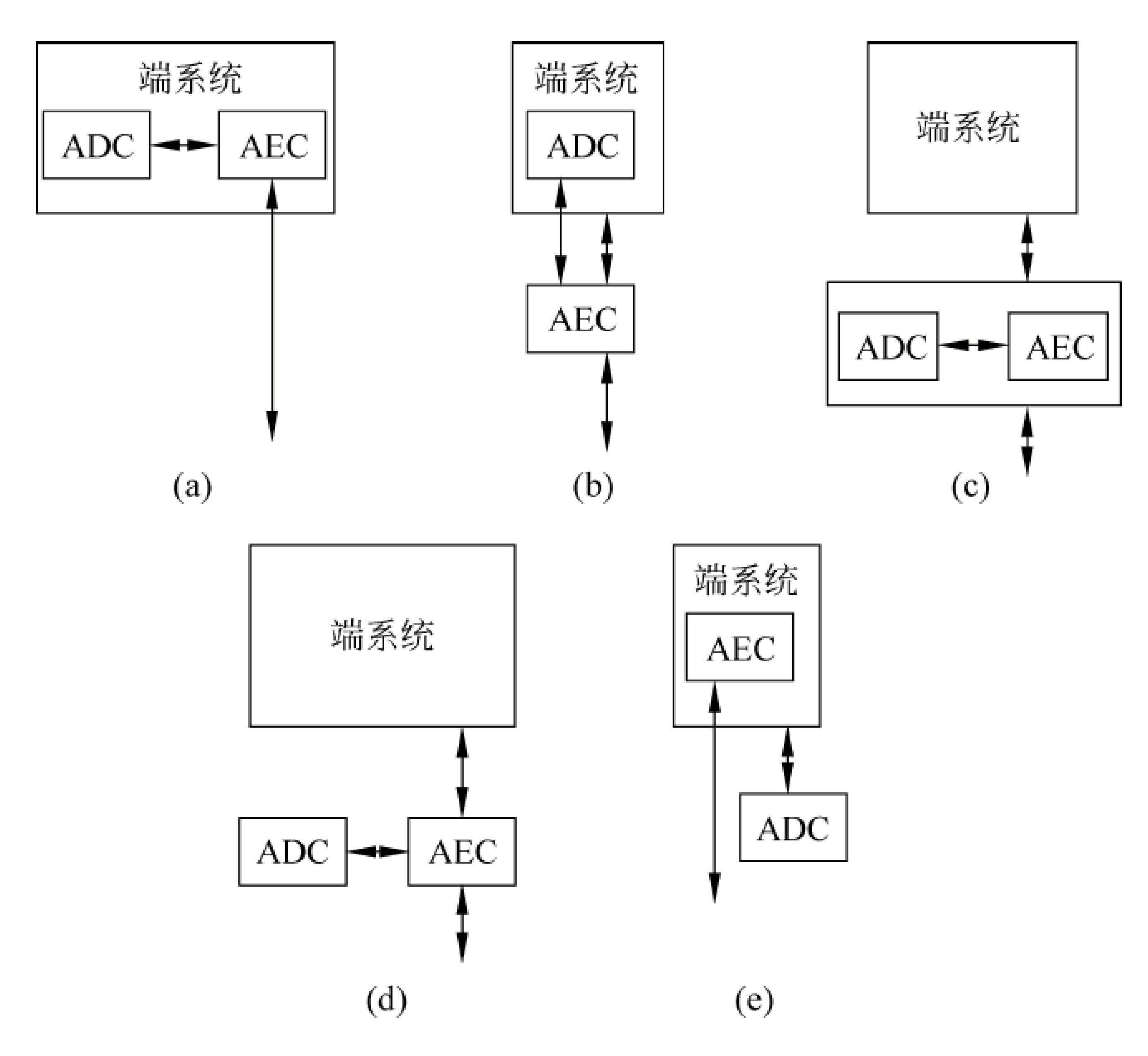


图 4-2 ADC/AEC 的 5 种不同的组合配置方式

(1) 在图 4-2(a)中,将 AEC 和 ADC 部署于端系统中,并授权其进行访问控制的查核工作,访问管理权限由端系统负责。

(2) 在图 4-2(b)中,将 AEC 独立于外部的服务器,端系统内的 ADC 根据外部 AEC 所传输来的执行要求来做执行权限的核查验证。端系统的系统管理者对其所属的信息资源有绝大部分的访问管理权限。

(3) 在图 4-2(c)中,将 AEC、ADC 两个功能组件独立于外部,但同属于一个可信赖的服务器所管辖。端系统必须根据外部的 ADC 所做出的决策判断来开放所属信息资源的使用权限。

(4) 在图 4-2(d)中,将 ADC、AEC 两个功能组件独立并分属两个可信赖的服务器管辖,访问控制的许可要同时经过两个以上的服务器判断后才能实施。

(5) 在图 4-2(e)中,将 ADC 独立于外部的服务器,端系统接收了发起者传输来的执行要求后,由外部的 ADC 来决策判断,并将查核结果通知端系统中的 AEC。端系统的系统管理者对其所属的信息资源有部分的访问管辖权。

若 AEC 和 ADC 采取紧密连接的配置方式时,由于同在一个组件内因而减缓了通信的延迟,可以大幅增加访问控制的时效性,而且可以避免 ADF 与 AEF 通信之间提供安全防护的负担。若是一个 ADC 支持多个 AEC 时,这种情况就像在分布式环境下,授权管制信息以集中方式存储在 ADC 中,各端系统除了不必存储访问控制的信息外,也可以减少权限审查组件的设置。

ADF 能由一个或多个 ADF 组件实现,且 AEF 能由一个或多个 AEF 组件实现。图 4-3

给出了访问控制组件间的关系示例。这里描述的关系只适用于单个发起者和单个目标，其他示例可能包括使用多于一个 ADC 和 AEC。

(1) 在图 4-3(a)中，发起者直接向发起者的 AEC 提交它请求的访问，要求 ADC 批准访问请求。若访问被批准，AEC 通报给请求的目标。

(2) 在图 4-3(b)中，发起者向目标的 AEC 直接提交请求的访问，随后 AEC 将它提交给 ADC 批准。若访问被批准，AEC 通报给请求的目标。

(3) 图 4-3 的(a)和(b)在功能和位置上相互对应。AEC 形成出或入访问控制，或者两者都形成，因此，AEC 可被称为发起者 AEC，或目标 AEC，或被插入的 AEC。

(4) 在图 4-3(c)中，发起者将请求的访问提交给插入的 AEC，随后 AEC 将它提交给 ADC 批准。若访问被批准，AEC 通报给请求的目标。

(5) 在图 4-3(d)中，交互是(a)图和(b)图与同一 ADC 的合成，而该 ADC 批准发起者和目标 AEC 的访问请求。发起者将其请求的访问提交给发起者的 AEC，AEC 请求 ADC 批准。若访问被批准，发起者的 AEC 将此请求的访问出示给目标的 AEC，随后该 AEC 将其出示给 ADC 批准。若请求被批准，则 AEC 通报给请求的目标。

(6) 图 4-3 的(e)和(f)中，分离的 AEC 强制实施出或入访问控制。在(e)图中，除了双方 AEC 都必须批准请求的访问外，交互与(c)图相似。在(f)图中，交互是(a)图与(b)图的合成，但使用分离的 AEC。

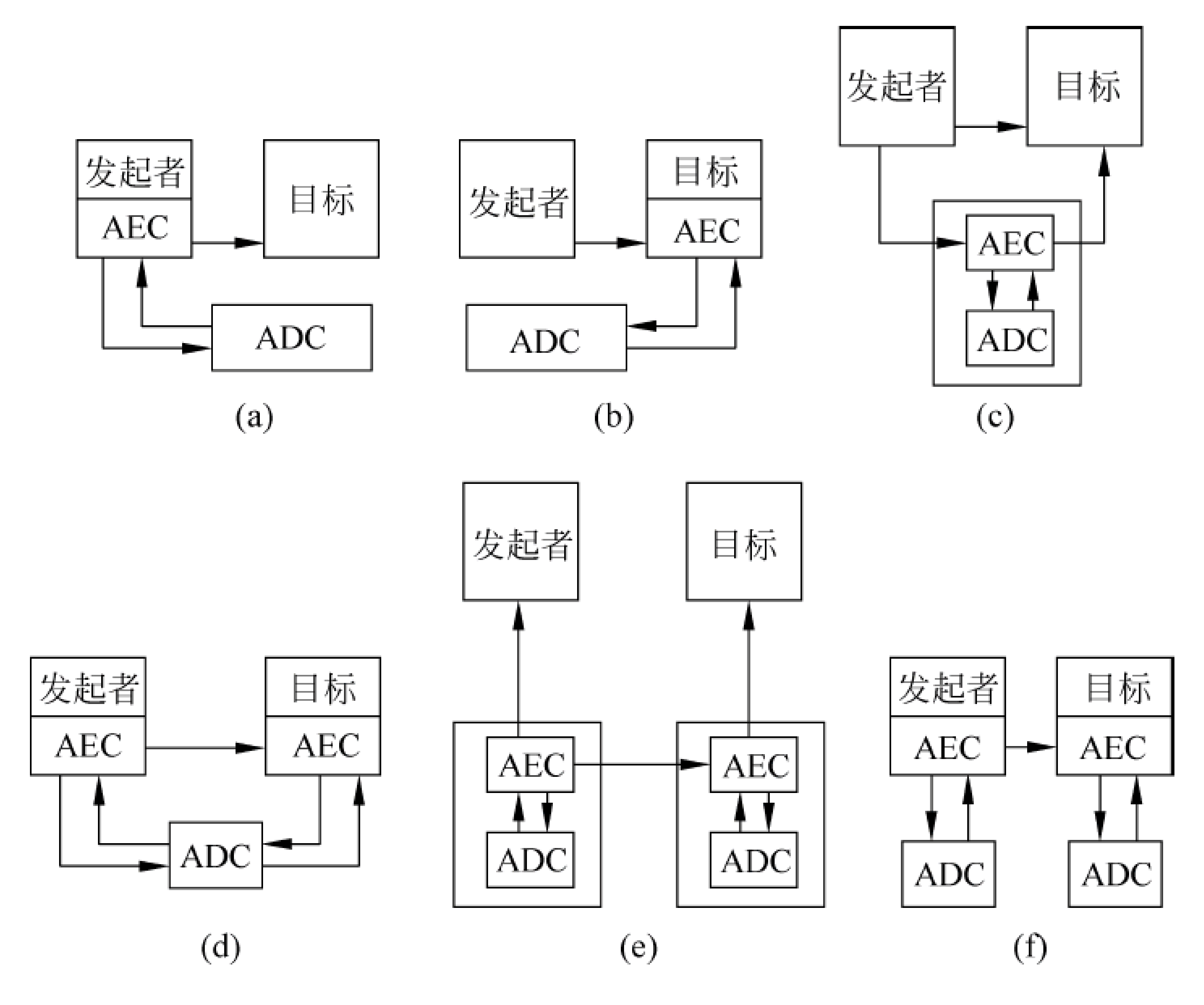


图 4-3　组件关系

4.1.4　访问控制活动

1. 建立访问控制策略的表示

通常用自然语言将访问控制策略陈述为概括性的原则，例如，只允许某一级别以上的管理人员才能检查员工的个人信息。将这些原则转换成规则是一项工程设计活动，必须在其他访问控制活动之前进行，但它不属于本安全框架的范畴。

2. 建立访问控制信息 ACI 的表示

在这项活动中，要对实时系统中的 ACI 表示和系统之间的交换做出选择。ACI 的表示必须能够支持特定访问控制策略的需求。不同的 ACI 表示可用于不同的目的以及用于特定的元素中间。

建立 ACI 表示的一个方面就是决定可被指定给安全域中元素的 ACI 值的类型和范围。经选择的 ACI 表示可看出模板，为安全域中的元素赋予特定的 ACI 值。

3. 给发起者和目标分配 ACI

在这一活动中，分配给一个元素的 ACI 具体属性类型和属性值是由安全域机构 SDA（Security Domain Authority）、SDA 代理或其他实体（例如资源拥有者）指定的。这些实体可根据安全域策略指定或修改 ACI 的分配。由一个实体分配的 ACI 可通过由另一个实体已经绑定到它的 ACI 加以限制。当有新元素添加到安全域中时，给元素分配 ACI 是一个不间断的活动。

ACI 可以是关于单一实体的信息，也可以是关于实体间关系的信息。分配给一个发起者或目标的 ACI 可以包括发起者 ACI、目标 ACI 或上下文信息。

在实际操作中，ACI 必须被绑定到一个元素上，使得一个采用从绑定 ACI 导出 ADI 的 ADF 信任那条信息。因此，尽管给元素分配 ACI 对构造绑定 ACI 而言是先决条件，但只有绑定到一个元素的 ACI 才能出现在实开放系统中。

4. 绑定 ACI 到发起者、目标和访问请求

将 ACI 绑定到一个元素会在元素和分配给元素的 ACI 之间创建一个安全链接。可能有几种绑定机制，其中有些依赖于元素和 ACI 的位置，而另一些可能依靠一些密码签名或封印处理。在某些安全策略下还需要维护 ACI 的机密性。

当有新元素添加到安全域中时，对元素的 ACI 绑定是一种不间断的活动。一个 SDA 的代理或其他实体可根据适用的安全策略任意删除或添加 ACI 绑定，在需要表达对安全策略或属性的变更时，SDA 可对绑定到元素的 ACI 进行修改。绑定 ACI 可包括有效期指示器，从而使以后可能撤销的 ACI 量减到最少。

5. 使 ADI 对 ADF 可用

如果访问控制策略允许且使用绑定机制认证的话，可由发起者或目标选择一个绑定到发起者或目标的 ACI 子集，在 ADF 中进行特定访问控制判决时使用。绑定到一个元素的 ACI 可暂时绑定到另一个元素。

对发起者 ADI、目标 ADI 或访问请求 ADI，存在以下 3 种可能：

(1) 在分配 ACI 值后，ADI 可被预置到一个或多个 ADF 组件中。

(2) ADI 可由在访问控制进程中递交给 ADF 组件的绑定 ACI 导出。

(3) ADI 可由从其他来源所获得的绑定 ACI 导出。根据需要，发起者或目标获得绑定 ACI，或者由 ADF 获得绑定 ACI。ADF 必须能够明确地确定 ADI 已由适当的 SDA 从绑定到元素的 ACI 导出。

6. 修改 ACI

SDI 可根据需要修改已分配并绑定到一个元素的 ACI，以表示变化中的安全属性。

ACI 可在分配给元素后的任何时间被修改。如果修改降低了发起者对目标访问的可允许度，则这种变更有可能要求撤销该 ACI 以及由其导出而可被 ADF 保留的 ADI。

7. 撤销 ADI

撤销 ACI 后，任何试图使用该 ACI 导出的 ADI 必然引起访问不被接受。在撤销 ACI 之前，应防止进一步使用由 ACI 导出的 ADI，否则使用它必然引起访问遭到拒绝。当撤销 ACI 时如果基于以前导出 ADI 的一个访问正在继续，那么，正在其中起作用的访问控制策略有可能要求终止该访问。

8. ACI 转发

在分布式系统中，常见的需求是一些实体请求其他实体代表它们去执行访问。发起者和目标是由实体所承担的角色，尽管并不是所有实体都可以承担这两个角色。一个实体在它本身担当另一个作为发起者的实体的目标时，还可同时承担对一个实体的发起者。

ACI 转发如图 4-4 所示，图中显示出实体 A 请求实体 B 对另一个实体 C 实施访问的基本概念。

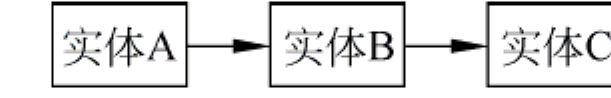


图 4-4　ACI 转发

这一基本概念有许多变形。这些变形在策略所要求的 ACI 的组合上有着明显不同，这些策略必须存在，允许进行这些链接访问，并表明如何使该 ACI 对合适的访问控制组件可用。在一些策略下，除为了实体 A 而执行访问已将 ACI 绑定到实体 B 以外，实体 B 可以不需要 ACI；在另一些策略下，实体 B 将仅使用从实体 A 得到的与该访问相关的 ACI，而在一般情况下，必须使用绑定到实体 A 和实体 B 的 ACI。

图 4-4 可推广到任何数量的具有最终目标实体 AEF 的中间实体，这些 AEF 主要以从序列中的一个或多个实体中获得的 ACI 为依据做出访问判断。

4.1.5　访问控制与其他安全措施的关系

目前，信息安全受到前所未有的挑战，单一的安全服务机制很难保证系统的真正安全，与其他安全技术结合将成为访问控制技术的趋势之一。这些安全服务机制包括认证、数据完整性、数据机密性、安全审计。

1. 身份认证

身份认证是最基本的措施，也是信息安全保密防范的第一道线，目的是验证通信双方的真实身份，防止非法用户假冒合法用户窃取敏感数据。在安全的通信中，涉及的通信各方必须通过某种形式的身份认证机制来证明身份，验证用户的身份与所宣称的是否一致，然后才能实现对于不同用户的访问控制和记录。

身份认证与访问控制尽管有某些共性和相互联系，但服务是不相同的。身份认证是判断访问者是否具有其声称的身份的处理过程。访问控制是信息安全保密防范的第二道防线，它是在身份认证成功的基础上取得用户身份，根据身份信息和授权数据库决定是否能够访问某个资源。

在某些系统中用于身份认证的验证设施与 ADF 搭配在一起。在分布式系统中无须搭配这些功能，并且可使用分离的发起者 ACI，因此身份信息仅被简单地当作发起者绑定 ACI

的一部分。

2. 数据完整性

数据完整性服务用于确保在访问控制组件内或组件之间输入和输出的完整性,例如防止 ACI 以及存储或传输中的上下文信息被修改。

3. 数据机密性

在一些安全策略控制下,可要求数据机密性服务以便对访问控制组件或访问控制组件之间的某些输入和某些输出实现机密性,例如防止收集敏感信息。

4. 安全审计

安全审计技术是一种事后追查手段。审计系统根据审计设置记录用户的请求和行为,同时入侵检测系统实时或非实时地检测是否有入侵行为。访问控制和审计系统都要依赖于身份认证系统提供的用户身份信息。

ACI 可用来审计一个特定发起者的访问请求,这需要收集若干审计线索,以便能够准确地识别哪个发起者执行了哪些访问请求。

审计策略可以要求将某些或所有访问企图记录下来,因此要求有一个用于访问控制机制的可靠记录机制。访问控制策略可以要求不进行审计就不能进行访问,在这种情况下访问控制机制将在功能上依赖于可靠的记录服务。

在要求发起者具有可确认性的情况下,发起者总是在访问前受到身份鉴别。身份和访问控制尽管经常紧密相关,但并不总是由同一机构控制下的功能来执行,也无须搭配这些功能。用于身份认证的信息可能需要用来获取发起者绑定 ACI。

总之,访问控制的目的是确保合法使用者对系统信息的使用权限。与身份认证机制的配合,使访问控制在实施上更有效;与数据完整性配合可确保执行的要求在授权管理功能单元之间传输时不会被非法使用者任意修改;而与数据机密性机制配合,可以确保授权管理信息的机密性,也就是在传输过程中不会被他人获取或破解,即使有黑客侵入授权管理信息库,也无法获得授权管理信息。

4.1.6 访问控制颗粒和容度

访问控制策略可在不同的颗粒级别上定义目标。每一个颗粒级别可有它自己的逻辑上的分离策略,并可限度使用不同 AEF 与 ADF 组件。例如,对一个数据库服务器的访问可被控制为该服务器仅作为一个整体的访问,要么完全拒绝发起者访问,要么允许访问服务器上的任何东西。另一种选择是,访问可控制到对单个文件、文件中的记录甚至是记录中的数据项。特定的数据库可以是目录信息树,对其访问控制粒度可以在整个树一级、数内的子树、树的条目甚至是条目的属性值。

通过规定一种策略,容度可用来对一个目标集实施访问控制,只有在对包含这些目标的一个目标被允许访问时,该策略才允许访问这些目标。容度也应用在包含于大组里的发起者子组。容度概念常常应用在互相关联的目标中,例如数据库中的文件或记录中的数据项。当一个元素被包含在另一个元素之中时,在试图访问经密封的元素之前,有必要给发起者赋予通过该密封元素的访问权力。

4.1.7　多级安全与访问控制

随着计算机网络特别是 Internet 的发展，多级安全的分布式应用成为一个研究重点。由于分布式应用的安全最终通过对用户之间消息的加密完成，所以加密传输的消息还得满足相应的存取控制策略，从而需要相应的密钥管理技术。对于类似问题的研究始于20世纪80年代，但是由于当时应用背景的限制，早期的研究主要针对数据库和操作系统。随着应用需求的不断增加，关于分布式应用中的多级安全存取控制策略的研究成为研究重点，但目前相关问题的发布体制的缺点如下：首先，所依靠的安全策略使得分组管理与密钥管理脱离；其次，主要依靠计算不同安全类别之间的关系参数实现，这种实现降低了效率，常用的策略是把不同用户分成不同的安全级别，然后根据所属安全级别进行密钥发布。

多级安全的概念始于20世纪60年代，当时美国国防部（DOD）决定开发一些保护计算机中存储的机密数据的措施。在这之前，一直是采用一些规章制度来限制那些未经批准人员处理系统内的机密数据，因为那时还没有值得信任的计算机能够有效地保护机内的机密数据。在多级安全系统中，所有信息都有一个密级，每个用户也都相应地有一个签证。要决定是否允许某用户读一个文件，就比较该用户的签证是否与该文件的密级相符。安全策略要求，为了合法地得到某一信息，用户的安全级必须大于或等于该信息的安全级，并且该信息属于用户的信息访问类别。随着网络技术的不断发展，信息资源的共享已很普遍，在多用户计算环境下如何管理信息资源愈来愈显得重要，多级安全体系中通过授权进行访问控制的技术可以直接应用于数据库中的信息管理和网络以及操作系统中的信息管理。利用密码技术实现多级安全系统中的访问控制也是一个重要的研究方向。

计算机系统内的用户可以按安全权限分成不相交的用户类集合：$U=\{U_1,U_2,\cdots,U_n\}$，每个用户都有一个相应的安全级。用偏序关系“$\leqslant$”表示 U 中用户间安全级的高低，$U_i\leqslant U_j$ 表示用户类 U_i 的安全级别不高于用户类 U_j 的安全级，从而“$U_i\leqslant$”构成一个偏序集。设 X_m 为一段信息或一个客体，授权中心希望把它存储在系统中或在系统中传播，下标 m 的含义是该客体可被用户类 U_m 访问，U 中的偏序关系蕴涵了对于所有满足 $U_m\leqslant U_i$ 的用户类 U_i 都可以访问客体 X_m。多级安全策略要求设计一个系统来满足上诉要求，下面给出利用密码技术实现上述要求的一种方法。假设存在一个已知的密码算法，加密函数 E，加密密钥为 K_e，解密函数为 D，解密密钥为 K_d，加、解密过程为：$C=EK_e(P)$，$P=DK_d(C)$。如果是对称密码体制，那么 K_e 和 K_d 完全相同。如果是非对称密码体制，那么 K_e 和 K_d 不同，并且不能由 K_e 和 C 公开信息计算出 K_d。利用密码技术实现多级安全的步骤如下：

（1）授权中心利用特定密码算法生成 n 个解密密钥 K_1、K_2、…、K_n。

（2）对 $i=1$、2、…、n，将 K_i 分配给用户 U_i，并由各用户秘密保存。

（3）对 $i=1$、2、…、n，所有满足 $U_i\leqslant U_j$ 的用户 U_j 都可以得到 K_i。

当一个客体 X_m 在系统中存储或传播时，首先利用 K_m 对它进行加密 $X'_m=EK_m(X_m)$，然后将[X'_m，m]存储或传播，只有拥有 K_m 的主体才能从 X'_m 中恢复出 $X_m=DK_m(X'_m)$。这种方法的优点是仅需存储或传播一个 X_m 的副本，加、解密也只进行一次。缺点是用户必须保存大量密钥，所以能够有效应用的关键是如何进行密钥管理。

4.2 访问控制的分类

传统的访问控制方法包括了美国国防部颁布的橘皮书(Orange Book)中制订的两种方法：强制的访问控制(Mandatory Access Control,MAC)和自由裁决的访问控制(Discretionary Access Control,DAC)。MAC使用安全标签(Security Label)控制信息的流向，而安全标签是由系统管理者强制指定的。此种方法适用于多阶层(Multi-level)的组织结构。其制订严格而缺乏弹性，无法适应于复杂的企业环境中。而在DAC方法下，资源的拥有者可以自行决定资源的访问权限，因其具有弹性的特性，目前已广泛应用于商用系统中，但其具有无法控制信息流向的缺点。在企业环境中，系统产生的资源是属于公司的财产，若由使用者自行裁决资源的使用权限，并非合理的作法。同时以上两种方法皆注重于最基本的访问权，诸如读(read)、写(write)、删除(revoke)、增加(append)、执行(execute)等，并无法满足复杂的企业环境需求。例如在一笔银行的交易中，可能就牵涉到数次读、写、执行的动作，若使用MAC或DAC来制订系统的访问政策(policy)，显然是件非常复杂的工作。

基于以上理由，美国国家标准局(NIST)提出了以角色为基础的访问控制(Role-Based Access Control,RBAC)方法。不同于DAC的是，在RBAC中，使用者皆被分配到适当的角色，而资源的访问权限则是经由所属的角色来决定，即使用者可以自行裁决所拥有资源的访问权限。不同于MAC与DAC偏重于作业系统或档案的访问控制，RBAC可视为应用层(Application Level)的访问控制，在使用者访问权限的描述上可以用较高层次的方式来描述，例如一个银行柜员的权限包括了存款、提款等。下面将详细介绍几种不同的访问控制方法。

4.2.1 强制访问控制(MAC)

强制访问控制(MAC)是指系统强制主体服从事先制定的访问控制策略。MAC在20世纪80年代得到普遍应用，主要用于多层次安全级别的军事应用中，在1985年美国国防部的TCSEC中被用作为B级安全系统的主要评价标准之一。

在MAC系统中，所有主体和客体都被分配了安全标签以标识一个安全等级。当主体访问客体时，调用强制访问控制机制，根据主体的安全等级和访问方式，比较主体的安全等级和客体的安全等级，从而确定是否允许主体对客体的访问。

计算机应用的发展促使研究人员设计并实现了众多的强制访问控制策略，目前已实现的比较成熟的策略有CAP(能力策略)、MLS(多级安全策略)、DTE(域及类型实施策略)、TE(类型实施策略)、FF(文件标识策略)和IBAC(基于标识的访问控制)等。

强制访问控制是对自主访问控制的重要补充，能有效地防止自主访问中存在的安全问题。因为在强制访问下，系统独立于用户行为强制执行访问控制，访问控制检查的依据是主体和客体的安全属性标记，最终是否允许访问是由系统实现的各安全策略决定，任何用户都无法改变控制规则。访问控制规则不会随着系统的运行、时间的推移而改变和扩散。例如某个主体无权访问某个文件，无论系统运行多久，无论主体执行什么操作，即使该文件被转

存,他始终无法访问该文件的内容,除非安全管理员提升用户的权限或修改文件的安全属性。因此强制访问控制实现了比自主访问控制更加严格的访问控制措施。受强制访问控制保护时,用户不会因为疏忽导致信息泄露,也不会产生访问权限扩散。即使存在特洛伊木马(Trojan Horse),强制访问控制也会保证信息在安全等级中按一个方向流动,解决了自主访问控制的问题。

首先,由于 MAC 增加了不能回避的访问控制,可能影响系统的灵活性,导致 MAC 策略的应用领域比较窄,一般只用于军方等具有明显等级观念且安全性要求高的行业或领域。其次,虽然 MAC 增强了信息的机密性,但它重点强调信息向高安全级的方向流动,对高安全级信息的完整性保护强调不够,而一些完整性控制策略却可以实现机密性的功能。最后,在 MAC 系统中实现单向信息流的前提是系统中不存在逆向潜信道。逆向潜信道的存在会导致信息反规则的流动,而现在计算机系统中这种潜信道是难以去除的,例如大量的共享存储器以及提供硬性性能而采用的各种 Cache 等,这给系统增加了安全性漏洞。

4.2.2　自主访问控制(DAC)

自主访问控制(DAC)最早出现在 20 世纪 60 年代末期的分时系统中,它是在确认主体身份及所属组的基础上,对访问进行限定的一种控制策略。在这种方式下,主体可以按照自己的意愿精确指定系统中的其他主体对其客体的访问权。其实现理论基础是访问控制矩阵(Access Control Matrix),它将系统的安全状态描述为一个矩阵,矩阵的行表示系统中的主体,列表示系统中的客体,每个元素表示主体对客体所拥有的访问权限。访问控制矩阵模型如表 4-1 所示。

表 4-1　访问矩阵模型

	Object1	**Object 2**	……	**Object *n***	……
Subject 1	read,write	read	……		……
Subject 2	read	write	……	read,write	……
……	……	……	……	……	……
Subject *m*	own,read,write	write	……	read	……
……	……	……	……	……	……

为了提高效率,系统不保存整个矩阵,在具体实现时是基于矩阵的行或列来实现访问控制策略的。目前自主访问控制主要由两种实现方式:

1. 基于行(主体)的 DAC 实现

通过在每个主体上都附加一个该主体可以访问的客体的明细表来表现,根据表中信息的不同可分为以下 3 种形式:

(1) 权能表(Capability):决定用户是否可以对客体进行访问以及进行何种形式的访问(读、写、修改、执行等)。一个拥有某种权力的主体可以按一定方式访问客体,并且在进程运行期间访问权限可以是添加或删除等。

(2) 前缀表(Profiles):包括受保护的客体名以及主体对它的访问权。主体欲访问某客体时,自主访问控制系统将检查主体的前缀是否具有它所请求的访问权。

(3) 密码(Password):每个客体(甚至客体的每种访问模式)都需要有一个密码,主体

访问客体时首先向操作系统提供该客体的密码。

2. 基于列(客体)的 DAC 实现

通过对每个客体附加一个可访问它的明细表来表现,有以下两种形式。

(1) 保护位(Protection Bits):保护位是对所有的主体指明一个访问模式集合,由于它不能完备地表达访问控制矩阵,因而很少使用。

(2) 访问控制列表(Access Control List,ACL):是目前采用得最多的方法,通过在客体上附加一个主体明细的方法表示访问控制矩阵,表中的每一项包括主体的身份和对客体的访问权,当删除一个主体时,要检查所有客体的 ACL。

ACL 的优点在于表述直观、易于理解,而且比较容易查出对一特定资源拥有访问的所有用户,有效地实施授权管理,可以在各种不同类型的系统应用中使用。但应用到规模较大、需求复杂的信息系统中时,客体与主体都非常多,ACL 将占据很大的存储空间,访问判决时系统也将占用很多 CPU 时间。且当主体发生变化时,例如当用户的职位、职责发生变化时,管理员要花费大量时间修改与主体相关的 ACL,这不但使得访问控制的授权管理花费很大的人力,而且也很容易出错。主体对客体的访问权限关系在信息移动过程中会被改变,不仅会造成对数据的无意泄露,也难以抵抗对数据的恶意攻击。

DAC 在一定程度上实现了多用户的权限隔离和资源保护,并且实现简便,通常应用于商业环境中。实际应用中大多数系统采用基于自主访问控制机制来实现访问控制,例如主流操作系统、防火墙等。但 DAC 的一个致命弱点是:访问权的授予是可以传递的。一旦访问权限被传递出去将难以控制,访问权的管理是很困难的,会带来严重的安全问题。另外,DAC 不保护受保护的可以产生的副本,即一个用户不能访问某一客体,但能够访问它的复制,这更增加了管理难度;最严重的是不能对系统中信息进行保护,信息容易泄露,无法抵御特洛伊木马的攻击。另外,在大型系统中无论使用哪一种形式的 DAC,系统开销都会很大,效率低,难以满足大型应用特别是网络应用的需要。

4.2.3 基于角色的访问控制(RBAC)

基于角色的访问控制(RBAC)模型是目前比较流行和先进的安全访问控制模型。RBAC 的概念开始于 20 世纪 70 年代的多用户和多机系统中,1992 年,D. Ferraiolo 和 R. Kuhn 在已有 RBAC 概念的基础上,形式化地定义了 RBAC 模型。RBAC 模型通过引入角色的概念,将访问控制中的主体对象和对应权限解耦。RBAC 是目前公认的解决大型企业的统一资源访问控制的有效访问方法,其具备的两个特征是:

(1) 由于角色/权限之间的变化比角色/用户关系之间的变化相对要慢得多,从而减小授权管理的复杂性,降低管理开销。

(2) 可以灵活地支持企业的安全策略,并对企业变化有很大的伸缩性。

1. RBAC 的相关概念

(1) 角色(Role):指一个组织或任务中的工作或位置,代表了一种资格、权利和责任。

(2) 用户(User):指一个可以独立访问计算机系统中的数据或数据表示的其他资源的主体。

(3) 权限(Permission):表示对系统中的客体进行特定模式的访问操作,这与实现的机

制密切相关。

2. RBAC 的模型

RBAC 模型可以分为 4 种类型，分别是基本模型 RBAC0(Core RBAC)、角色分级模型 RBAC1(Hierarchal RBAC)、角色限制模型 RBAC2(Constraint RBAC)和统一模型 RBAC3(Combines RBAC)。RBAC 基本模型包含了 RBAC 标准最基本的内容，该模型的定义如图 4-5 所示。

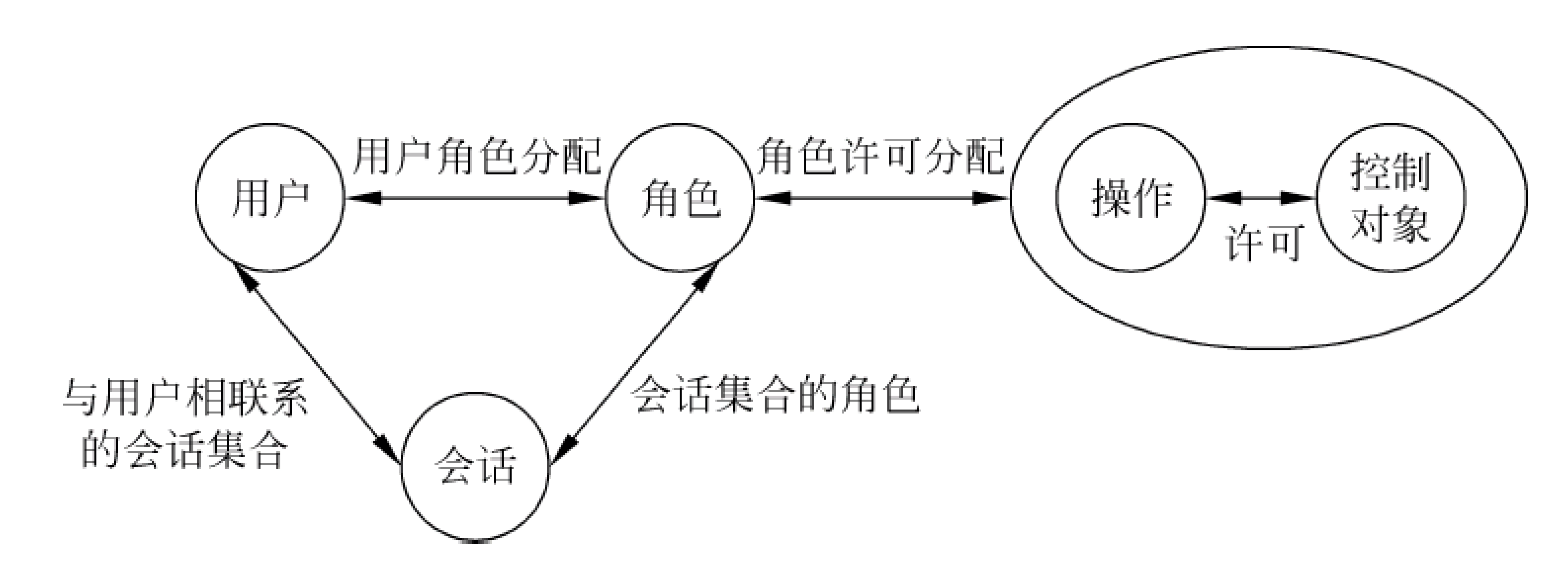


图 4-5　RBAC 模型

(1) 用户集(Users)：系统中的主体，执行操作。

(2) 对象集(Objects)：系统中被动的实体，主要包括被保护的信息资源。

(3) 权限集(Permissions)：每个元素分别来自于 Objects 和 Operations 的两个元素，即对象上的操作构成了权限。

(4) 角色集(Roles)：核心，联系用户与特权。

(5) 会话集(Sessions)：系统登录或通信进程和系统之间的会话。

它们之间的关系可表示为：用户被分配一定角色，角色被分配一定许可权，会话是用户与激活的角色集合之间的映射，而用户与角色间的关系定义和角色与权限间的关系定义无关。

RBAC 的核心思想是将权限与角色联系起来，在系统中根据应用的需要为不同的工作岗位创建相应的角色，同时根据用户职责指派合适的角色，用户通过所指派的角色获得相应的权限，实现对文件的访问。也就是说，传统的访问控制是直接将访问主体(发出访问操作，有存取要求的主动方)和客体(被调用的程序或欲存取的数据访问)相联系，而 RBAC 在中间加入角色，通过角色沟通主体和客体。这种访问控制方式能够灵活地表达和实现组织的安全策略，接近日常生活的应用需求。

4.2.4　基于任务的访问控制(TBAC)

1993 年，R. K. Thomas 等人首先提出了基于任务的访问控制(Task-Based Access Control，TBAC)思想，1994 年给出了基于任务的授权模型的概念基础，并于 1997 年明确提出了基于任务的访问控制模型。所谓任务就是用户要进行的一个个操作的统称。任务是一个动态的概念，每项任务包括其内容、状态(例如静止态、活动态、等待态、完成态等)、执行结果、生命周期等。任务与任务之间一般存在相互关联，这种关联可以相互依赖，也可以相互排斥。

TBAC 从任务角度进行授权控制，在任务执行前授予权限，在任务完成后收回权限。

在 TBAC 中，访问权限是与任务绑定在一起的，权限的生命周期随着任务的执行被激活，并且对象的权限随着执行任务的上下文环境发生变化，当任务完成后权限的生命周期也就结束了，因此它属于一种主动安全模型。TBAC 的一些相关概念如下所示。

(1) 授权步(Authorization Step)：是指在一个工作流程中对处理对象(例如办公流程中的原文档)的一次处理过程。它是访问控制所能控制的最小单元。授权步由受托人集(Trustee-set)和多个许可集(Permissions-set)组成。

(2) 授权结构体(Authorization Unit)：是由一个或多个授权步组成的结构体。它们在逻辑上是联系在一起的。授权结构体分为一般授权结构体和原子授权结构体。一般授权结构体内的授权步依次执行，原子授权结构体内部的每个授权步紧密联系。其中任何一个授权步失败都会导致整个结构体的失败。

(3) 任务(Task)：是工作流程中的一个逻辑单元。它是一个可区分的动作，可能与多个用户相关，也可能包括几个子任务。在实际工作中，一个任务包含如下特征：长期存在；可能包括多个子任务；完成一个子任务可能需要不同的人。

(4) 依赖(Dependency)：是指授权步之间或授权结构体之间的相互关系，包括顺序依赖、失败依赖、分权依赖和代理依赖。依赖反映了基于任务的访问控制的原则。

尽管 TBAC 具备许多特点，并已应用于实际中，但当其应用于复杂的企业环境时，自身的一些缺陷就逐渐显露出来。例如在实际的企业环境中，角色是一个非常重要的概念，但 TBAC 中并没有将角色与任务清楚地分离开来，也不支持角色的层次等级；另外，TBAC 并不支持被动访问控制，需要与 RBAC 结合使用。

4.2.5 其他访问控制方式

在访问控制技术的发展过程中。为了适应不同的应用需求，研究学者们不断提出不同的访问控制模型。

1. 基于角色和任务的访问控制模型

1998 年，G. Coulouris 等人在 RBAC 和 TBAC 模型的基础上，提出了基于角色和任务的访问控制模型(Task-Role-Based Access Control，T-RBAC)，它将 RBAC 和 TBAC 结合起来，把任务置于角色和权限之间，给用户指派角色，再给任务分配角色，同时规定执行任务时需要的最小访问权限，这样，当用户提出访问请示时，通过拥有角色来获得某个任务的相关访问权限。该模型继承了 RBAC 和 TBAC 模型的优点，非常适合应用在工作流管理系统中。

2. 基于规则策略的访问控制模型

E. Bertino 等人在 RBAC 模型的基础上给出了一个基于规则的授权模型，该模型提出一种约束描述语言，它既能表达静态约束，也能表达动态约束，并且给出了约束规则的一致性检查算法。朱羚等人也提出了一种基于约束规则的访问控制模型(Constraint-Based Access Control，CBAC)，该模型采用显式授权与隐式授权相结合的安全机制，引进一种用于形式化语言来精确描述 CBAC 模型安全策略，并制定了一种描述用户属性约束和时间属性约束的统一语法规范。

3. 面向服务的访问控制模型

面向服务的访问控制模型是最近几年才发展起来的。随着数据库、网络和分布式计算机的发展，组织任务进一步自动化，与服务相关的信息进一步计算机化，增加了工作流访问控制的复杂性。研究人员从工作流访问控制模型与流程模型分离角度来解决此问题。中国科学院软件研究所的徐伟等人提出了一种面向服务的工作流访问控制模型，该模型中服务是流程任务的抽象执行和实施访问控制的基本单元，通过服务将组织角色、流程任务和执行权限关联起来，避免了访问控制模型与流程模型的直接关联。

4. 基于状态的访问控制模型

2001 年，B. Steinmuller 等人将 RBAC 模型扩展，提出了一个基于状态的 RBAC 扩展模型。该模型在传统 RBAC 模型的基础上引入了状态的概念，将由对象访问控制的变化所引起的 RBAC 组件的变化作为状态的迁移，这样就可以为每个对象的访问控制构造一个状态转换图，从而可以根据状态转换图来跟踪各个对象的访问控制策略。该模型中的状态概念跟工作流运行中的任务状态和过程状态的概念非常类似，因此可以将其应用于工作流系统中。

5. 基于行为的访问控制模型

李凤华等人提出了一种基于行为的访问控制模型（Action-Based Access Control Model，ABAC），模型中的行为综合了角色、时间状态和环境状态的相关安全信息。ABAC 模型不仅可以提供传统的角色、角色控制和时态约束，还提供环境约束，支持移动计算的接入用户，接入的具体业务需求、接入位置、接入时间和接入平台具有随机性、事先不可预知等典型特性。因此，ABAC 具有广泛的应用范围、方便的应用方式。

4.3 AAA 技术

4.3.1 AAA 技术概述

随着信息技术的不断发展与完善，虚拟专用网（VPN）、远程拨号等移动接入的应用更为广泛。在新的网络应用环境下，传统用户身份认证和访问控制机制早已不能满足企业和用户的需求，AAA 认证授权机制应运而生。AAA 包含 3 个方面的内容：认证（Authentication）、授权（Authorization）和审计（Accounting），现在人们常常将它们称作“3A”。AAA 已成为网络安全策略研究的重要部分，并应用于各种网络的安全设计中，其主要目的是管理哪些用户可以访问网络服务器，具有访问权的用户可以得到哪些服务，如何对正在使用网络资源的用户进行审计。AAA 是提供给安全设备来授权用户接入设备或者临近网络的一个结构。授权特性用来在用户被认证之后限制用户的权限。审计用来维持设备活动，以及网络或者网络设备中作为日志。

AAA 可以在单个用户或者单个服务的基础上执行它的功能。换句话说，它可以用来认证和授权单个用户或者服务，例如 IP 和 IPX。这使 AAA 可以用于与它的 3 个功能相关的很多方面。

AAA 的工作相当简单。AAA 设置在路由器或者 PIX 或者任何其他这样的设备上，这

些设备都需要 AAA 对接入设备本身或者与设备相连的网络的用户进行限制。路由器可以使用本地数据获取用于 AAA 的数据,例如用户名或者密码或者每个用户的访问控制列表;也可以通过诸如 RADIUS(远程鉴权拨入用户服务)或者 TACACS+(终端访问控制器访问控制系统)这样的协议来请求一个认证服务器。AAA 模型允许认证、授权或者统计功能的执行独立于使用的协议(例如 TACACS+、RADIUS、Kerberos 等)。

1. 认证

认证用于识别用户在允许远程登录访问网络资源之前对其身份进行识别。整个认证通常是采用用户输入用户名与密码来进行权限审核。认证的原理是每个用户都有一个唯一的权限获得标准。由 AAA 服务器将用户的标准同数据库中每个用户的标准一一核对。如果符合,那么该用户认证通过,如果不符合则拒绝提供网络连接。如果设立了授权参数的话,就必须依据授权参数进行访问和使用。

在小范围内的认证通常可以使用路由器或者 PIX 或其他这类进行认证的设备上维持的一个密码列表来完成。但是,对于大范围内的认证,通常希望将接入设备(例如路由器或者 PIX)进行密码认证的工作分给一个专用的服务器,例如 TACACS+或 RADIUS 服务器。接入设备将用户名认证参数,例如接收来自进行认证的用户或者设备的用户名和密码,传递给服务器。然后服务器认证用户和密码是否跟数据库中的数据匹配。这些服务器可以接纳更高级的认证方法,例如一次性(One-time)密码、可变密码和基于外部数据库(例如 Windows NT 或者 UNIX 数据库)的认证。

2. 授权

授权是用户或设备被给予对网络资源受控制的访问权限的过程。授权让网络管理员控制谁能够在网络中做些什么。它也可以用来做这样的一些工作,例如给通过 PPP 服务连接的用户赋予指定的 IP 地址,要求用户使用特定类型的服务进行连接,或者配置 callback 之类的高级特性。

当启动 AAA 授权时,网络接入服务器使用从用户配置文件检索到的信息来配置用户的会话,这些信息要么位于本地用户数据库,要么位于安全服务器上。完成这个工作之后,如果用户设置文件的信息允许的话,用户就会被授予访问特定服务器的权限。

3. 审计

审计是 AAA 的最后一个内容。审计是网络接入服务器用来报告认证的和/或授权的用户及设备所做行为进行统计的过程,AAA 服务器通过 RADIUS 或者 TACACS+进行统计。鉴于认证和授权对用户和设备访问网络资源进行了限制,审计就负责进一步的工作并记录被认证和/或被授权的用户的行为。另外,审计也可以用来跟踪接入设备状态和 TACACS+或 RADIUS 通信。

审计信息以统计记录的形式在接入设备和 TACACS+或 RADIUS 服务器之间进行交换。每个统计记录包含统计属性-值对,并存储在 TACACS+或 RADIUS 服务器上。网络管理、客户统计和审计都可以利用这些数据进行分析。

4.3.2 AAA 协议

AAA 由具体的 AAA 协议来实现,目前最常用的 AAA 协议包含 RADIUS 和

TACACS＋两种。众多厂商都支持 AAA 协议，例如，微软内置的 Internet 身份认证服务(IAS)可以支持 RADIUS，Cisco ACS 可支持 RADIUS 和 TACACS＋协议。

1. 远程鉴权拨入用户服务(RADIUS)

远程鉴权拨入用户服务(Remote Authentication Dial In User Service，RADIUS)主要用于管理远程用户的网络接入行为。RADIUS 基于客户机/服务器(C/S)结构，其客户端最初就是 NAS 服务器(Net Access Server)，现在只要是运行 RADIUS 客户端软件的计算机都可称为 RADIUS 的客户端。RADIUS 协议认证机制较为灵活，可以采用 PAP、CHAP 或者 UNIX 登录认证等多种方式。RADIUS 协议中规定了网络接入服务器与 RADIUS 服务器之间的消息格式。RADIUS 整个运行模式为：服务器接收用户的连接请求，根据其账户和密码完成验证后，把用户所需的配置信息返回给网络接入服务器，RADIUS 服务器同时根据用户的动作进行审计并记录其计费信息。

2. 终端访问控制器访问控制系统(TACACS＋)

终端访问控制器访问控制系统(Terminal Access Controller Access Control System，TACACS＋)是 Cisco 为适应不断增长的安全市场的需求所研发出来的。它是一个全新的协议，采用可靠的应用传输控制协议(TCP)进行传输。RADIUS 从用户角度结合了认证和授权这两个部分，而 TACACS＋分离了这两个操作。RADIUS 和 TACACS＋的主要差异如表 4-2 所示。

表 4-2　RADIUS 和 TACACS＋比较

AAA 协议支持	RAIDUS	TACACS＋
端口号	认证/授权 1645/1812 审计 1646/1813	49
传输层协议	UDP	TCP
加密方法	仅对密码字段加密	对整个数据包加密
认证和授权是否分离	否	是
互操作性	基本不能互相操作	Cisco 设备支持全部特性
应用标准	工业标准	Cisco 专有

4.4　VPN 概述

4.4.1　VPN 的基本概念

VPN(Virtual Private Network，虚拟专用网)是指利用密码技术和访问控制技术在公共网络中建立的专用通信网络。在虚拟专用网中，任意两个节点之间的连接并没有传统专用网所需的端到端的物理链路，而是利用某种公众网的资源动态组成，虚拟专用网络对用户端透明，用户好像使用一条专用线路进行通信。

IETF 草案理解基于 IP 的 VPN 为："使用 IP 机制仿真出一个私有的广域网"，指的是通过私有的隧道技术在公共数据网络上仿真一条点到点的专线技术。所谓虚拟，是指用户不再需要拥有实际的长途数据线路，而是使用 Internet 公众数据网络的长途数据线路。所

谓专用网络,是指用户可以为自己制定一个最符合自己需求的网络。

虚拟专用网是网络互联技术和通信需求迅猛发展的产物。Internet 技术的快速发展及其应用领域的不断推广,使得许多部门(例如政府、外交、军队、跨国公司)越来越多地考虑利用廉价的公用基础通信设施构建自己的专用广域网络,进行本部门数据的安全传输,它们客观上促进了 VPN 在理论研究和实现技术上的发展。

4.4.2 VPN 的技术要求

实际应用中,虽然各 VPN 供应商可以采取多种不同的实现技术,但一个高效、成功的 VPN 必须满足以下基本要求。

1. 安全保障

所有的 VPN 均应保证通过公用网络平台传输数据的专用性和安全性。在安全性方面,由于 VPN 直接构建在公用网上,实现简单、方便、灵活,但同时其安全问题也更为突出。VPN 用户必须确保其传输的数据不被攻击者窥视和篡改,并且要防止非法用户对网络资源或私有信息的访问。VPN 可以利用加密技术对经过隧道传输的数据进行加密,以保证数据仅被指定的发送者和接收者了解,从而保证数据的私有性和安全性。

2. 服务质量(QoS)保证

不同的用户和业务对服务质量(Quality of Service,QoS)保证的要求差别较大,VPN 应当为他们提供不同等级的服务质量保证。例如,对于移动办公用户,提供广泛的连接和覆盖性是保证 VPN 服务的一个主要因素;而对于拥有众多分支机构的专线 VPN 网络,交互式的内部企业网应用则要求网络能提供良好的稳定性。在网络优化方面,QoS 通过流量预测与流量控制策略可以按照优先级分配带宽资源,实现带宽管理,使得各类数据能够被合理地先后发送,并预防阻塞的发生。

3. 可扩展性和灵活性

VPN 必须能够支持通过 Intranet(内联网)和 Extranet(外联网)的任何类型的数据流,方便增加新的节点,支持多种类型的传输媒介,可以满足同时传输语音、图像和数据等应用对高质量传输以及带宽增加的需求。

4. 可管理性

VPN 的管理主要包括安全管理、设备管理、配置管理、访问控制列表管理及 QoS 管理等内容。VPN 用户虽然可以将一些次要的网络管理任务交给服务提供商去完成,但自己仍需要完成许多网络管理任务。所以,一个完善的 VPN 管理系统是必不可少的。VPN 管理的目标为减小网络风险、高扩展性、经济性、高可靠性等。

从 VPN 的技术要求,可以看出其相对于传统专用网具有明显的优势,具体体现在以下 3 个方面:

(1) 可以降低成本。VPN 利用现有 Internet 或其他公共网络的基础设施为用户创建安全隧道,无须使用专门的线路,例如 DDN(数字数据网)和 FR(帧中继),只需要接入当地的 ISP(Internet 服务提供商)就可以安全地接入内部网络,这样就节省了线路费用,降低了成本。

(2) 可扩展性强。如果直接采用专线连接,在分布增多、内部网络节点越来越多时,网

络结构趋于复杂,费用也越来越昂贵。如果采用 VPN 技术,只需在节点处架设 VPN 设备,就可以利用 Internet 建立安全连接。

(3) 提供安全保证。VPN 技术利用可靠的加密认证技术,在内部网络之间建立隧道,能够保证通信数据的机密性和完整性,保证信息不被泄漏或暴露给未授权的实体并确保信息不被篡改。

4.4.3　VPN 的类型

根据 VPN 的应用环境,通常可以把 VPN 分成 3 种类型:远程访问虚拟网(Access VPN);企业内部虚拟网(Intranet VPN);企业扩展虚拟网(Extranet VPN)。

1. 远程访问虚拟网(Access VPN)

该类型的 VPN 主要用来处理可移动用户、远程交换和小部门远程访问企业本部的连通性。当出差人员需要和企业或相关部门联系时,便可以利用本地相应的软件接入 Internet,通过 Internet 和企业网络中相关的 VPN 网关建立一条安全通道。用户使用这条可以提供不同级别的加密和完整性保护的通道,可以传输不同级别保护的信息,但前提条件是用户所在地必须具备提供相应 VPN 功能的软件。如果用户所在地没有这些软件,只要 ISP 的接入设备可以提供 VPN 服务的话,用户也可以拨入 ISP,由 ISP 提供的 VPN 设备和企业本部的 VPN 网关进行安全通道的连接,并提供相似的安全数据传输服务。

2. 企业内部虚拟网(Intranet VPN)

企业内部虚拟网主要是利用 Internet 来连接企业的远程部门。在传统的企业内部网络的实现中,通常是采用专线方式来连接企业和各个远程部门的,这样需要为每一个远程部门申请一条专线,其运行、维护和管理费用之高是不言而喻的。与此同时,在这样的线路上传输的数据量通常比较少,很多时候带宽都得不到有效利用。而 VPN 恰好解决了这一问题,它只需企业的远程部门通过公用网络和企业本部互联,并且由远程部门网络的 VPN 网关和企业本部网络的 VPN 网关负责建立安全通道,在保证数据的机密性、完整性的同时又能大大地降低整个企业网互联的运行和管理费用。

3. 企业扩展虚拟网(Extranet VPN)

该类型网络主要用来连接相关企业和客户的网络。和 Intranet VPN 类似,传统的实现方案中,主要也存在费用较高,需要进行复杂的配置等诸多不便。而 VPN 可以在一定程度上解决这些问题,通过与 Internet 的互联,在降低了整个网络运行费用的同时又能在其他软件的辅助下较好地进行用户访问控制与管理。

4.4.4　VPN 的安全技术

VPN 的安全技术是其所有技术中最为关键的技术。目前,VPN 主要采用 4 项技术来保证安全,这 4 项技术分别是隧道技术(Tunneling)、加/解密技术(Encryption Decryption)、密钥管理技术(Key Management)和身份认证技术(Authentication)。其中隧道技术是整个 VPN 技术的核心。

1. 隧道技术

隧道技术是 VPN 的基本技术,类似于点对点连接技术,它在公用网建立一条数据通道

(隧道),让数据包通过这条隧道传输。隧道实质上是一种封装,它把一种协议 A 封装在另一种协议 B 中传输,实现协议 A 对公用网络的透明性。隧道根据相应的隧道协议来创建。

隧道可以按照隧道发起点位置划分为自愿隧道(Voluntary Tunnel)和强制隧道(Compulsory Tunnel)。自愿隧道由用户或客户端计算机通过发送 VPN 请求进行配置和创建,此时用户端计算机作为隧道客户方成为隧道的一个端点。强制隧道由支持 VPN 的拨号接入服务器配置和创建,此时用户端的计算机不作为隧道端点,而是由位于客户计算机和隧道服务器之间的远程接入服务器作为隧道客户端,成为隧道的一个端点。隧道技术在 VPN 的实现中具有如下主要作用:

(1) 一个 IP 隧道可以调整任何形式的有效负载,使远程用户能够透明地拨号上网来访问企业的 IP、IPX 或 AppleTalk 网络。

(2) 隧道能够利用封装技术同时调整多个用户或多个不同形式的有效负载。

(3) 使用隧道技术访问企业网时,企业网不会向 Internet 报告它的 IP 网络地址。

(4) 隧道技术允许接收者滤掉或报告个人的隧道连接。

2. 加/解密技术

加密技术是数据通信中一项较成熟的技术。利用加密技术保证传输数据的安全是 VPN 安全技术的核心。为了适应 VPN 工作特点,目前 VPN 中均采用对称加密体制和公钥加密体制相结合的方法。

对称加密体制(也称常规加密体制)的通信双方共享一个秘密密钥,发送方使用该密钥将明文加密成密文,接收方使用相同的密钥将密文还原成明文。对称加密算法运算速度快,因而 VPN 中将其用于加密要传输的数据,为了加大保密强度和便于程序实现,实际运用中多采用分组密码算法。

VPN 目前常用的对称密码加密算法有 DES、3DES、RC4、RC5、IDEA 和 CAST 等。

公钥加密体制,或称非对称加密体制,是通信各方使用两个不同的密钥,一个只有发送方知道的秘密密钥,另一个则是与之对应的公开密钥,公开密钥不需要保密。在通信过程中,发送方用接收方的公开密钥加密消息,并且用发送方的秘密密钥对消息的某一部分或全部加密,进行数字签名。接收方收到消息后,用自己的秘密密钥解密消息,并使用发送方的公开密钥解密数字签名,验证发送方身份。当前常见的公钥体制有 RSA、D-H 和椭圆曲线等,相应的加密算法都已应用于 VPN 实际实现中。

3. 密钥管理技术

密钥管理技术的主要任务是如何实现在公用数据网上安全地传递密钥而不被窃取。现行密钥管理技术分为 SKIP 与 ISAKMP/OAKLEY 两种。

SKIP 是由 SUN 公司提出的 Internet 简单密钥管理协议,它基于一个 D-H 公钥密码体制数字证书。SKIP 隐含地在通信双方实现了一个 D-H 交换,它简单易行,对公钥操作次数少,节省了系统资源,但由于公钥长期暴露,因而存在着安全隐患。

ISAKMP 是由美国 NSA 提出的 Internet 安全关联和密钥管理协议,是一个建立和管理安全关联(SA)的总体框架。它定义了默认的交换类型、通用的载荷格式、通信实体间的身份鉴别机制以及安全关联的管理等内容。OAKLEY 协议实际上提出了一种密钥生成方案,通过这种方案,可以使经过认证的通信双方利用 D-H 密钥交换方法,来协商产生安全的

秘密密钥材料。而 SKEME 协议则是一种能提供匿名性、可否认性的密钥生成方案。

ISAKMP/OAKLEY 协议(又称 IKE),即通常所说的 Internet 密钥交换协议。它综合了 OAKLEY 和 SKEME 的优点,使用了 ISAKMP 的语言,规范和综合了 OAKLEY 和 SKEME 的密钥交换方案,形成了一套具体的验证加密材料生成技术,以协商共享的安全策略。

4. 身份认证技术

VPN 中最常用的身份认证技术是用户名/密码或智能卡认证等方式。

身份认证是通信双方建立 VPN 的第一步,保证用户名/密码特别是用户密码的机密性至关重要。在 VPN 实现上,除了强制要求用户选择安全密码外,还特别采用对用户密码数据加密存放或使用一次性密码等技术。智能卡认证具有更强的安全性,它可以将用户的各种身份信息及公钥证书信息等集中在一张卡片上进行认证,做到智能卡的物理安全就可以在很大程度上保证认证机制的安全。

4.5　VPN 隧道协议

VPN 具体实现是采用隧道技术,而隧道是通过隧道协议实现的,隧道协议规定了隧道的建立、维护和删除规则以及怎样将企业网的数据封装在隧道中进行传输。隧道协议可分为第二层(链路层)隧道协议 PPTP、L2F、L2TP 和第三层(网络层)隧道协议 GRE 和 IPSec,如图 4-6 所示。

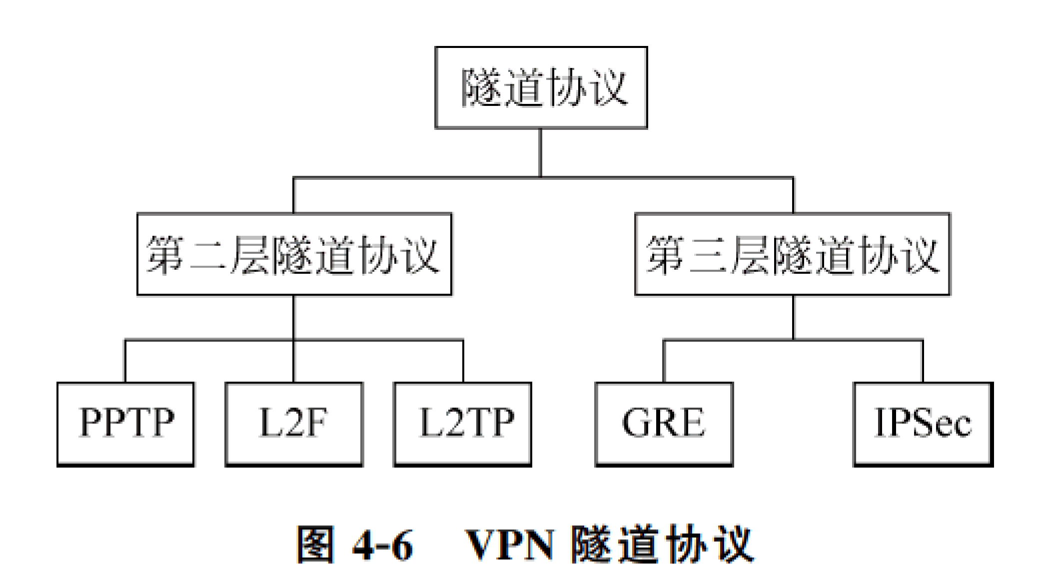


图 4-6　VPN 隧道协议

4.5.1　第二层隧道协议

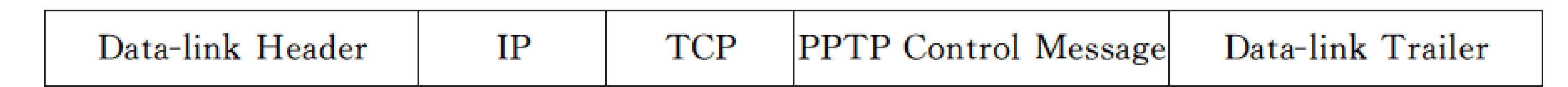
1. PPTP

PPTP(Point-to-Point Tunneling Protocol,点对点隧道协议)是在 PPP(Point-to-Point Protocol,点对点协议)的基础上开发的一种新的增强型隧道协议。利用 PPP 协议的身份认证、加密和协议配置机制,PPTP 为远程访问和 VPN 连接提供了一条安全路径。PPTP 通过控制连接来创建、维护和终止一条隧道,并使用 GRE(Generic Routing Encapsulation,通用路由封装)对经过加密、压缩处理的 PPP 帧进行封装。通过 PPTP,用户可以采用拨号方式接入到公共网络。PPTP 通信主要由 PPTP 控制连接和 PPTP 数据隧道两部分组成。

1) PPTP 控制连接

PPTP 的控制连接是一种必须通过一系列 PPTP 消息来创建、维护与终止的逻辑连接。PPTP 控制连接通信过程使用 PPTP 客户端上动态分配的 TCP 端口以及 PPTP 服务器上

编号为1723的反向IANA TCP端口。PPTP控制连接数据包包括一个IP包头、一个TCP包头和PPTP控制信息,具体如图4-7所示。



Data-link Header	IP	TCP	PPTP Control Message	Data-link Trailer

图4-7 PPTP控制连接数据包

PPTP控制连接的过程如下:

(1) 在PPTP客户端上动态分配的TCP端口与PPTP服务器上编号1723的TCP端口之间建立一条TCP连接。

(2) PPTP客户端发送一条用以建立PPTP控制连接的PPTP消息。

(3) PPTP服务器通过一条PPTP消息进行响应。

(4) PPTP客户端发送另一条PPTP消息,并且选择一个用以对从PPTP客户端向PPTP服务器发送数据的PPTP隧道进行标识的调用ID。

(5) PPTP服务器通过另一条PPTP消息进行应答,并且为自己选择一个用以对从PPTP服务器向PPTP客户端发送数据的PPTP隧道进行标识的调用ID。

(6) PPTP客户端发送一条PPTP Set-Link-Info消息,以便指定PPP协商选项。

2) PPTP数据隧道

当通过PPTP连接发送数据时,PPP帧将使用GRE报头进行封装,GRE报头包含了用以对数据包所使用的特定PPTP隧道进行标识的信息。

初始PPP有效载荷如IP数据包、IPX数据包或NetBEUI帧等经过加密后,添加PPP报头,封装形成PPP帧。PPP帧再进一步添加GRE报头,经过第二层封装形成GRE报文,在第三层封装时添加IP包头。IP包头包含数据包源地址及目的端IP地址。数据链路层封装是IP数据包多层封装的最后一层,依据不同的外发物理网络再添加相应的数据链路层报头和报尾。

PPTP数据包在接收端的处理过程如下:

(1) 处理并去除数据链路层包头和包尾。

(2) 处理并去除IP包头。

(3) 处理并去除GRE和PPP包头。

(4) 如果需要的话,对PPP有效载荷即传输数据进行解密或解压缩。

(5) 对传输数据直接接收或者转发处理。

2. L2F

L2F(Level 2 Forwarding Protocol,第二层转发协议)是由Cisco公司提出的可以在多种传输网络上建立多协议的一种隧道协议,当然它也采用了Tunneling技术,主要面向远程或拨号用户的使用。L2F可以在多种传输介质(例如ATM、FR)上建立VPN通信隧道。它可以将链路层协议封装起来进行传输,因此网络的链路层独立于用户的链路层协议。

L2F远程接入过程为:远程用户按照常规方式拨号到ISP的接入服务器NAS,建立PPP连接,NAS根据用户名等信息再发起第二重连接,呼叫用户网络的服务器。整个过程中,L2F隧道的建立和配置对于用户来说是完全透明的。L2F主要强调的是将物理层协议移到链路层,并允许通过Internet光缆的链路层和较高层协议的传输,物理层协议仍然保持

在对该 ISP 的拨号连接中。一旦建立连接，L2F 将通过在保持初始拨号服务器位置不可见的 Internet 中的虚拟隧道来传输包含验证、授权和审计信息的数据包。此外，L2F 还可解决 IP 地址和审计的问题，它对可靠地处理这两个问题提供建议并打下基础。

3. L2TP

L2TP(Layer 2 Tunneling Protocol，第二层隧道协议)是一种工业标准 Internet 隧道协议，它把链路层 PPP 帧封装在公共网络设施(例如 IP、ATM、FR)中进行隧道传输。L2TP 结合了 PPTP 协议以及 L2F 协议的优点，能以隧道方式使 PPP 数据包通过各种网络协议。与 PPTP 不同，L2TP 隧道的维护不在独立的连接上进行，数据信息的传输是通过多级封装实现的。在安全性上，L2TP 仅仅定义了控制包的加密传输方式，对传输中的数据并不加密。

L2TP 系统由认证模块、日志模块、LAC(L2TP Access Concentrator，L2TP 访问集中器)模块和 LNS(L2TP Network Server，L2TP 网络服务器)模块组成。其中 LAC 用于发起呼叫、接收呼叫和建立隧道，为用户提供网络接入服务，具有 PPP 端系统和 L2TP 协议处理能力。LNS 是用于处理 L2TP 协议服务器端部分的软件。认证、日志模块是共用模块，LAC 和 LNS 都需要使用，如图 4-8 所示。

1) 认证模块

认证模块有一个极为重要的数据资源——用户认证数据库，库中由多个用户信息记录组成。每个用户记录由用户号、用户组、用户真实姓名、用户认证协议、用户使能状态构成和 CHAP 共享秘密组成。当然，这里的用户对 LNS 而言是 LAC，对 LAC 而言是 LNS，认证服务的工作原理如图 4-9 所示。



图 4-8　L2TP 系统功能模块

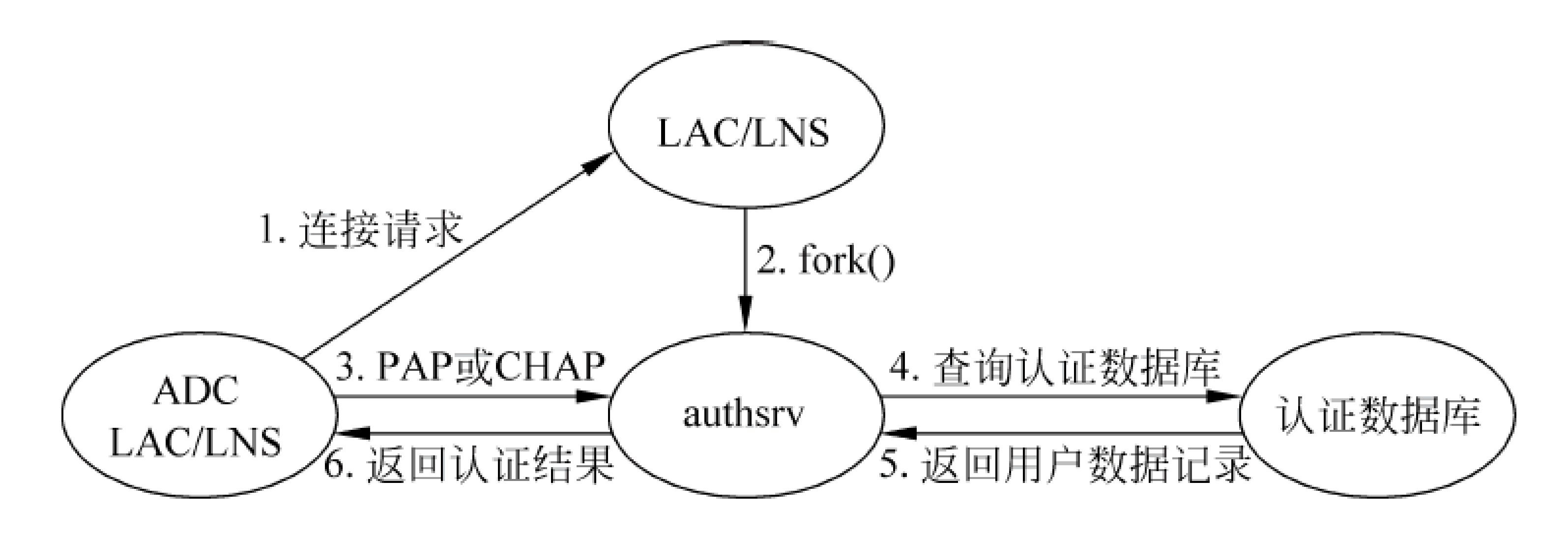


图 4-9　认证服务的工作原理

2) 日志模块

日志模块作为一个函数库使用，例如 WrSysLo()接口，日志功能是成熟系统的一大标志。在如此复杂的系统中，日志将在系统审计中起着重要作用。L2TP 系统不仅提供一般的系统日志，还对 L2TP 的包进行分类分级，例如系统日志、数据包(包括 PPP)日志和控制包日志都以独立的日志文件存在，在必要的时候，通过日志级别可审计不同的日志。

3) LAC 模块

当一个入站调用请求到达时(例如电话拨号)，将由 LAC 生成入站调用消息；检测 LNS 的连接，如果没有建立，则初始化到 LNS 的连接，构造滑动窗口队列的大小；生成新的出站控制包，加入控制消息，设置状态为 SCCRQ，生成挑战，并标示要挑战对方的位为真，但此时并不能预测响应，因为不知道对方的主机名；发出开始控制连接请求包(SCCRQ)，等待

开始控制连接响应包(SCCRP);当收到一个开始连接的响应,如果一切正常,根据 SCCRP 的主机名和给定的主机名计算挑战值,如果与 SCCRP 中的期望值一致,就发送开始控制连接包(SCCCN),否则发送停止控制连接包(stopCCN),清除隧道;等待 HELLO 包;在一定的时间延迟内,如果收到 HELLO 包,则创建成功,发出 ACK 包,否则创建控制连接失败,清除隧道。

4) LNS 模块

当收到一个 SCCRQ 的请求时,首先检查 SCCRQ 是否可以接收,如果是则生成一个新的控制连接响应包,生成挑战值,并在包中指明期望的响应值,发出 SCCRP 包,否则发出 stopCCN 包,清除隧道;等待接收 SCCCN 包,看是否可以接收,如果是则计算挑战值,如果与 SCCCN 中的期望值一致,发出 HELLO 包,等待;如果收到 ACK 包则表明控制连接创建成功。

L2TP 的建立过程是:

(1) 用户通过公共电话网或 ISDN 拨号至本地的接入服务器 LAC,LAC 接收呼叫并进行基本的辨别。

(2) 当用户被确认为合法企业用户时,就建立一个通向 LNS 的拨号 VPN 隧道。

(3) 企业内部的安全服务器(例如 RADIUS)鉴定拨号用户。

(4) LNS 与远程用户交换 PPP 信息,分配 IP 地址。LNS 可采用企业专用地址(未注册的 IP 地址)或服务提供商提供的地址空间分配 IP 地址。因为内部源 IP 地址与目的地 IP 地址实际上都通过服务提供商的 IP 网络在 PPP 信息包内传输,企业专用地址对提供者的网络是透明的。

(5) 端到端的数据从拨号用户传到 LNS。

在实际应用中,LAC 将拨号用户的 PPP 帧封装后传输到 LNS,LNS 去掉封装包头得到 PPP 帧,再去掉 PPP 帧头得到网络层数据包。

4.5.2 第三层隧道协议

1. GRE

GRE(Generic Routing Encapsulation,通用路由封装)是网络中通过隧道将通信从一个专用网络传输到另一个专用网络的常用到的一个协议,它属于网络层协议。

它的运行过程通常是这样的:当路由器接收了一个需要封装的上层协议数据报文,首先这个报文按照 GRE 协议的规则被封装在 GRE 协议报文中,而后再交给 IP 层,由 IP 层再封装成 IP 协议报文便于网络的传输,等到达对端的 GRE 协议处理网关时,按照相反的过程处理,就可以得到所需的上层协议的数据报文了。

GRE 具有如下优点:多协议的本地网可以通过单一协议的骨干网实现传输;可以将一些不能连续的子网连接起来,用于组建 VPN;扩大了网络的工作范围,包括那些路由网关有限的协议,例如 IPX 包最多可转发 16 次,而在一个隧道连接中看上去只经过一个路由器。

图 4-10 所示为使用 GRE 来封装分组的一般形式。

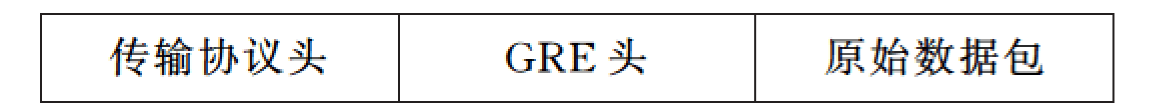


图 4-10 GRE 分组格式

传输协议头是 IPv4 的头。有效载荷分组可以是 IPv4 的头，或者其他协议。GRE 允许非 IP 协议在有效载荷中传输。使用 IPv4 头的 GRE 分组被归入 IP 协议，类型号为 47。当为 GRE 生成过滤时这是一条很重要的信息。当 GRE 中封装的分组是 IPv4 时，GRE 头协议类型域被设定为 0x800。

在 GRE 领域中有两个主要实现：一个基于 RFC1701；另一个基于较新的 RFC2784，它也是推荐的协议标准。RFC2784 的实现与 RFC1701 的实现在某种程度上有互用的部分，但 RFC2784 中摒弃了在 RFC1701 中提供的一些特性。图 4-11 所示为基于 RFC1701 的 GRE 头格式。

<table>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>C</td><td>R</td><td>K</td><td>S</td><td>s</td><td colspan="3">Recurl</td><td colspan="5">Flags</td><td colspan="3">Ver</td><td colspan="16">Protocol Type</td></tr>
<tr><td colspan="16">Checksum(Optional)</td><td colspan="16">Offset(Optional)</td></tr>
<tr><td colspan="32">Key(Optional)</td></tr>
<tr><td colspan="32">Sequence Number(Optional)</td></tr>
<tr><td colspan="32">Routing(Optional)</td></tr>
</table>

图 4-11　基于 RFC1701 的 GRE 头格式

图中重要的域说明如下：

（1）**标识符**（以下描述是从 RFC1701 中摘录的，同时加以说明）：

校验和存在位 C（第 0 位）：当该位被置为 1 时，在分组中存在校验和域同时该域包含的信息有效。如果校验和存在位或路由存在位置位，则在 GRE 分组中给出校验和与偏移域。

路由存在位 R（第 1 位）：当该位被置为 1 时，分组中给出偏移域和路由域并且包含有效信息。如果校验和存在位或路由存在位被置位，则在 GRE 分组中给出校验与偏移域。

密钥存在位 K（第 2 位）：当该位被置为 1 时，在 GRE 头中给出密钥域。否则在 GRE 头中不给出密钥域。

序号存在位 S（第 3 位）：当该位被置为 1 时，分组存在序号域，否则在 GRE 头中不给出序号域。

严格源路由位 S（第 4 位）：建议该位当且仅当路由信息由严格源路由组成时才置为 1。

递归控制 Recurl（第 5～7 位）：递归控制是一个 3 位的无符号整数，该整数包含被允许附加封装的数目。它在默认情况下应当被置为 0。

标识位 Flags（第 8～12 位）：在 RFC1701 中没有定义。

版本号 Ver（第 13～15 位）：版本号域必须包含数值 0。

（2）**协议类型**（Protocol Type，2B）：协议类型域存放封装在 GRE 分组的有效载荷中的分组的协议类型。例如，当 IP 是 GRE 分组中运载的协议时，本域设置为 0x800，当运载协议为 Novell IP X 时，其值为 Ox8137。

偏移量（Offset，2B）：本域指出 Routing 域到净荷的字节偏移。

（3）**校验和**（Checksum，2B）：本域被用来保证 GRE 头和有效载荷中的完整性。它存放一个 GRE 头和有效载荷分组的 IP 校验和。

(4) **密钥**(Key,4B):密钥域存在一个用来认证已封装 GRE 分组的数字。这是 GRE 提供的一种形式上较弱的安全性。基本上,密钥域防止了误配置或其他源地址的篡改。隧道的两端只接收密钥域正确的 GRE 分组。密钥域需要在隧道两端手动配置。很显然,安全性不能依赖于它,因为当攻击者只需要简单地查看 GRE 分组就可以算出密钥域的值,从而可以产生一个可认证的 GRE 分组,如同原始封转一样。密钥域的另一个用途是标识隧道中单独的通信流。例如,分组需要根据没有在封装的数据中出现的上下文信息来确定路由。密钥域提供了上下文信息并在封装与解封装之间定义了一个逻辑通信流。

(5) **序号**(Seguence Number,4B):网络两端可以使用序号跟踪接收到的分组顺序,并且可以选择性丢弃乱序到达的分组。这部分在传输通信协议是有用的,但是在接收乱序分组时效果较差(例如基于 LLC2 协议)。

(6) **路由**(Routiong,4B):路由域列出源路由入口(SRE)。此域使用的不是很频繁,只有需要对 GRE 分组作源路由时才用到它。

讨论完 RFC1701 的 GRE 实现,再看看 RFC2784 的实现。RFC2784 不赞同 GRE 中的 3 个可选域(序号、密钥和路由)并把它们和同在 RFC1701 中使用的校验和域一同除去。在这些标识符域的位置使用 0 做替代。如果分组是由 RFC 2784 发出的并被 RFC 1701 接收到,这样就可保证 RFC 2784 和 RFC 1701 实现之间的互用性。RFC1701 实现仅把 0 作为序号,密钥、路由选项没有使用。不过,如果发送器使用 RFC1701,并且在实现时将那些 RFC 2784 废弃掉的某个域置为 1,那么发出的分组必将被使用 RFC2784 实现的接收器丢弃。

2. IPSec

PPTP、L2F 和 L2TP 协议各自有自己的优点,但是都没有很好地解决隧道加密和数据加密的问题。而 IPSec(IP Security,IP 安全协议)协议把多种安全技术集合到一起,可以建立一个安全、可靠的隧道。这些安全技术包括:Diffie-Hellman 密钥交换技术;DES、RC4、IDEA 数据加密技术;哈希散列算法 HMAC、MD5、SHA;数字签名技术等。

1) IPSec 的定义

根据 IETF 标准,IPSec 的定义如下:网络层中的一个安全协议,为提供加密安全服务而开发,该服务可以灵活地支持认证、完整性、访问控制以及数据一致性。IPSec 是安全联网的长期方向,它通过端对端的安全性来提供主动的保护以防止专用网络与 Internet 的攻击。

2) IPSec 的安全结构

IPSec 安全结构包括 3 个基本协议:AH 协议为 IP 包提供信息源验证和完整性保证;ESP 协议提供加密保证;密钥管理协议(ISAKMP)提供双方交流时的共享安全信息。IPSec 通过上述 3 个基本协议在 IP 包头后增加新的字段来实现安全保证。图 4-12 所示为一个 IPsec 数据包的格式。

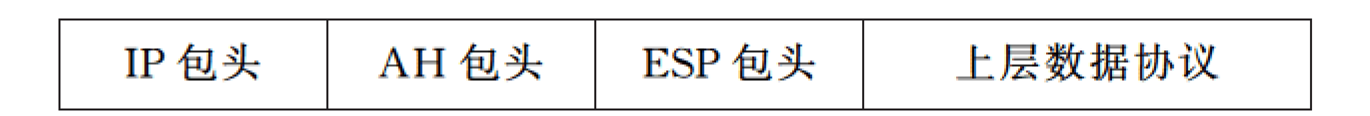


图 4-12 IPSec 数据包格式

3) IPSec 的工作方式

IPSec 采用两种工作方式:隧道模式和传输模式。

在隧道方式中，整个用户的 IP 数据包被用来计算 ESP 包头，整个 IP 包被加密并和 ESP 包头一起封装在一个新的 IP 包内。这样当数据在 Internet 上传输时，真正的源地址和目的地址被隐藏起来。

在传输模式中，只有高层协议(TCP、UDP、ICMP 等)及数据进行加密。在这种模式下，源地址、目的地址以及所有 IP 包头的内容都不加密。

4) IPSec 的分类

在大多数情况下，IPSec 允许在两个专用网络之间创建一个加密隧道。它同时允许隧道两端的认证。不过，IPSec 协议只是允许 IP 数据的封装和加密(GRE 可以隧道传输非 IP 流量，但不能对其加密)，所以如果为非 IP 流量创建隧道，IPSec 就得同诸如 GRE 一样的协议联合使用，它们允许隧道传输非 IP 协议。IPSec 试图解决的 VPN 两个主要设计问题：

(1) 把两个专用网络组合成一个虚拟网络的无缝连接。

(2) 将虚拟网络扩展成允许远程访问用户(也称为 Road Warriors)成为可信网络的一部分。

基于以上两个设计的基础，IPSec VPN 可以分为两大类：LAN-to-LAN IPSec 实现和远程访问客户端实现。

(1) LAN-to-LAN IPSec 实现。

LAN-to-LAN IPSec 描述的是在两个局域网之间建立的 IPSec 隧道的概念，也被称作 site-to-site VPN。建立 VPN-to-VPN 时，两个专用网络之间跨越一个公用网络，这样在任意一个专用网络中的用户都可以访问另一个专用网络中的资源，就像他们在各自的专用网络上一样。

(2) 远程访问客户端 IPSec 实现。

当一个远程用户连接到一个 IPSec 路由器或使用安装在其上的 IPSec 客户端访问服务器时，就会创建远程访问客户端 IPSec VPN。一般情况下，这些远程访问机器使用拨号或类似的连接方式连接到公用网络或 Internet。一旦到 Internet 的连接建立起来后，IPSec 客户端就可以建立一条跨越公共网络或 Internet 而连接到一个位于专用网络边缘的 IPSec 终端设备的封装隧道。远程访问客户端正是试图与这个专用网络建立连接并成为其中的一部分。这些 IPSec 终端设备也被称为一个 IPSec 远程访问集线器。

IPSec 的远程访问实现也有一些自己独有的挑战。在 LAN-to-LAN 情形中，IPSec 对等体的数量，即 IPSec 隧道的终端设备是有限制的。不过，在远程访问 IPSec VPN 的例子中，终端设备的数目是很多的，甚至可以达到成百上千。这些情形需要特殊的可扩展性好的认证密钥管理机制，因为替所有的用户保存所有密钥几乎是不可能的任务。

4.5.3　各种隧道协议比较

与 PPTP 和 L2F 相比，L2TP 的优点在于提供了差错和流量控制；L2TP 使用 UDP 封装和传输 PPP 帧。UDP 是一种非连接的传输协议，无法保证网络数据的可靠传输，L2TP 使用 N_r(下一个希望接收的消息序列号)和 N_s(当前发送的数据包序列号)字段控制流量和差错。双方通过序列号来确定数据包的次序和缓冲区，一旦数据丢失，根据序列号可以进行重发。L2TP 还定义了控制包的加密传输，每个被建立的隧道可以生成一个独一无二的随机钥匙，以便抵抗欺骗性的攻击，但是它对传输中的数据并不加密。

IPSec同其他隧道协议一样，不仅可以保证隧道的安全，同时还有一整套保证用户数据安全的措施，利用他建立起来的隧道更具有安全性和可靠性。IPSec还可以和L2TP、GRE等其他隧道协议一同使用，给用户提供更大的灵活性和可靠性。此外，IPSec可以运行于网络的任意一部分，它可以在路由器和防火墙之间、路由器和路由器之间、PC和服务器之间、PC和拨号访问设备之间运行，相当灵活方便。

从纵向来看，第三层隧道协议与第二层隧道协议相比更具有安全性、可扩展性及可靠性。从安全的角度来看，第二层隧道一般终止在用户网设备上，对用户网的安全及防火墙技术要求很高；而第三层的隧道一般终止在ISP的网关上，不会对用户网的安全构成威胁。从可扩展性角度来看，第二层隧道将整个PPP帧封装在报文内，可能会产生传输效率问题，PPP会话贯穿整个隧道，并终止在用户网的网关或服务器上，导致用户网内的网关要保存大量的PPP对话状态及信息，这会对系统负荷产生较大的影响，也影响系统的扩展性。除此之外，由于PPP的LCP(数据链路层控制)及NCP(网络层控制)对时间非常敏感，隧道的效率会造成PPP会话超时等问题；而第三层隧道终止在ISP网内，并且PPP会话终止在RAS(Remote Access Service，远程访问服务)，网点无须管理和维护每个PPP会话状态，从而减轻了系统负荷。

4.6 VPN的应用和发展趋势

4.6.1 VPN应用发展趋势

由于VPN技术复杂，协议多，用户需求千差万别，因此目前市场上存在不同类型和规格的VPN产品。按处理速度分，VPN产品可分为低端和高端两种；按应用平台则可分为软件平台、专用硬件平台和辅助硬件平台3种。国外一些大公司，例如3Com、Lucent、Cisco、Intel等都有成套的低-高端VPN产品。国内一些网络安全公司，例如清华同方、东大阿尔派等也都有自己的VPN产品。

未来几年里，VPN的需求市场有非常广阔的前景，VPN产品将进入高速增长期。未来80%～95%的企业将采用VPN实现宽带Internet。

4.6.2 VPN技术发展趋势

1. 基于IPSec的VPN产品将成为市场的主流

IPSec技术相对成熟，具有良好的安全机制，又得到IETF组织的推崇，因而受到越来越多厂商、用户的青睐。即使是PPTP的开创者——微软公司也逐渐倾向采用IPSec来开发VPN。基于IPSec的VPN产品将成为市场的主流产品。

2. VPN所用密码算法的抗攻击性不断增强

为了提高安全防护能力，采用56比特的DES或其他短密钥(例如40比特)密码算法的VPN越来越少，大部分VPN产品采用128比特以上长度的密钥(例如密钥长度为168比特的3DES)。

3. VPN 向集成化的方向发展

与安全路由器、防火墙等产品集成，是目前较为常见的 VPN 解决方案。将多种网络安全服务如隧道技术、IPSec、密钥交换技术、防火墙技术、QoS 与配置管理等集成于 VPN 产品，并支持功能扩展，为用户提供良好的选择和性价比将是未来 VPN 的一个发展方向。

4. 新的 VPN 实现技术将会不断推出

VPN 技术的研究仍是一个新的领域，是当今计算机网络研究的热点之一。在国际上，VPN 技术一方面向着标准化和应用化发展；另一方面，新的 VPN 实现技术必将不断推出，以满足各种不断增长的网络通信需求。

思　考　题

(1) 什么是访问控制技术？
(2) 访问控制有哪些功能及组件？
(3) 访问控制可以分为哪几类？简要介绍其优缺点。
(4) AAA 由哪几部分组成？每部分的具体功能有哪些？
(5) AAA 服务器是如何提供授权信息到 NAS 的？
(6) 什么是授权者请求域？
(7) 比较 RADIUS 和 TACACS+协议的差异。
(8) VPN 是如何定义的？
(9) VPN 可以分为哪几种类型？简要说明各种类型的应用环境。
(10) VPN 隧道协议可以分为哪几类？
(11) 请描述 L2TP 系统功能模块，并简要介绍各部分的功能。
(12) 简述 PPTP 控制连接的过程。
(13) 什么是 IPsec？IPsec 由哪 3 种主要的协议组成？

参 考 文 献

[1] 刘宏月，范九伦，马建峰. 访问控制技术研究进展. 小型微型计算机系统，2004，(1)：56～59.
[2] 陈颈. 访问控制技术的研究. 福建电脑，2005，(3)：11～12.
[3] Ravi S，Sandhu，Edward J. Coyne，Hal L. Feinstein，Charles E. Youman. Role-based access control models. IEEE Computer，1996.(2)：38～47.
[4] Ravi S. Sandhu. Access Control：The neglected frontier. ACISP，1996.
[5] 郭玮，茅兵，谢立. 强制访问控制 MAC 的设计及实现. 计算机应用与软件，2004，(3)：1～3.
[6] 单智勇，孙玉芳. 通用访问控制框架扩展研究. 计算机研究与发展. 2003，(2)：228～234.
[7] 沈海波，洪帆. 访问控制模型研究综述. 计算机应用研究，2005，(6).
[8] 尹绍锋. 访问控制技术研究及应用. 湖南大学硕士研究生学位论文，2008.
[9] Saadat Malik. 网络安全原理与实践. 北京：人民邮电出版社，2008.
[10] 罗鑫. 访问控制技术与模型研究. 北京邮电大学博士研究生学位论文，2009.

[11]　王悦.访问控制技术的研究与应用.天津财经大学硕士研究生学位论文,2007.
[12]　Mark Lucas,Abhishek Singh,Chris Cantrell.防火墙策略与 VPN 配置.北京：中国水利水电出版社，2008.
[13]　蒋东毅,吕述望,罗晓广.VPN 的关键技术研究.计算机工程与应用,2003,(15)：173～177.
[14]　魏广科.VPN 技术及其应用研究.计算机工程与设计,2005,(3)：714～715.
[15]　谢方军,戴宗坤,张红,等.VPN 中的分布式访问控制.小型微型计算机系统,2004,(7)：1250～1252.
[16]　赵阿群,吉逸,顾冠群.支持 VPN 隧道技术研究.通信学报,2000,(6)：85～91.
[17]　郝辉,钱华林.VPN 及其隧道技术研究.微电子学与计算机,2004,(11)：47～51.
[18]　过林吉,沈浅.VPN 隧道协议的研究和探讨.电脑知识与技术,2010,(6)：609～610.
[19]　舍堃,谭兴烈,周明天.L2TP 虚拟专用网.电子科技大学学报,2002,(4)：383～386.
[20]　贾铁军.网络安全管理及实用技术.北京：机械工业出版社,2010.

第 5 章　防火墙与入侵检测技术

本章学习目标

当前，很多企业已经将自己的内部网络和 Internet 连接，这样不仅可以便利地同商业伙伴开展业务，还可以充分利用 Internet 中广阔的资源。但是在获取资源的同时也带来了一些安全隐患，防火墙和入侵检测技术应运而生。

通过对本章的学习，应掌握以下内容：

(1) 网络安全的目的、意义及相关技术。

(2) 防火墙的基本概念和种类。

(3) 防火墙的体系结构及功能。

(4) 入侵检测技术的种类及各类技术的相关性能。

防火墙技术是应用广泛的网络安全技术，它通过监测、限制和更改跨越防火墙的数据流等多种技术，尽可能地对外部网络屏蔽有关受保护网络的结构信息。防火墙可以隔离风险区域和安全区域的连接，同时不会妨碍对风险区域的访问，还可以监控进出网络的通信量，预防不希望的、未授权的信息进出被保护的网络，筑起网络的第一道安全防线。

作为网络安全技术的重要一员，入侵检测技术已成为当今一种非常重要的动态安全技术，它与传统的静态安全技术相结合，达到了比较理想的安全目的。入侵检测技术的重点在于如何有效地提取攻击特征数据并准确地分析出不正常的入侵行为。

5.1　防火墙技术

5.1.1　防火墙的概念

防火墙是指隔离在本地网络与外界网络之间的一个执行访问控制策略的防御系统，是这一类防范措施的总称。在 Internet 上防火墙是一种非常有效的网络安全模型，通过它可以隔离风险区域与安全区域(局域网)的连接，同时不会妨碍用户对风险区域的访问。防火墙放在受保护网络与外部网络之间，如图 5-1 所示。

防火墙实质上是一种隔离控制技术，其核心思想是在不安全的网络环境下构造一种相对安全的内部网络环境。从逻辑上讲它既是分析器又是限制器，它要求所有进出网络的数据流都必须遵循安全策略，同时将内外网络在逻辑上分离。

防火墙能增强机构内部网络的安全性。防火墙系统决定了哪些内部服务可以被外界访问，外界的哪些人员可以访问内部的服务，哪些外部服务可以被内部人员访问。防火墙必须只允许授权的数据通过，而且防火墙本身也必须能够免于渗透。

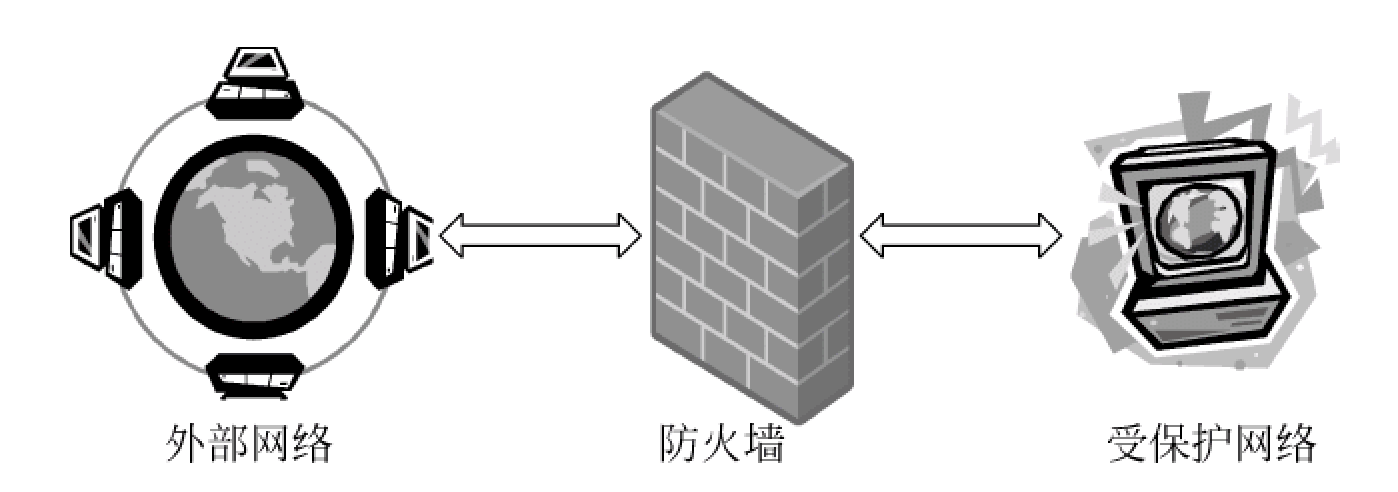


图 5-1 防火墙示意图

5.1.2 防火墙的种类

防火墙有许多种形式,有以软件形式运行在普通计算机之上的,也有以固件形式设计在路由器之中的。总的来说可以分为 3 种:包过滤防火墙,应用级网关防火墙,状态监测型防火墙。

1. 包过滤防火墙

在互联的 TCP/IP 网络上,所有往来的信息都被分割成许许多多一定长度的数据包,每一个数据包中都会包含一些特定信息,例如数据的源地址、目标地址、TCP/UDP 源端口和目标端口等。当这些数据包被送上互联网络时,路由器会读取接收者的 IP 并选择一条合适的物理线路发送出去,数据包可能经由不同的路线抵达目的地,当所有的包抵达目的地后会重新组装还原。包过滤型防火墙会检查所有通过的数据包中的 IP 地址,并按照系统管理员所给定的过滤规则进行过滤,一旦发现来自危险站点的数据包,防火墙便会将这些数据拒之门外。

包过滤技术的优点是它对于用户来说是透明的,处理速度快而且易于维护,实现成本较低,在应用环境比较简单的情况下,能够以较小的代价在一定程度上保证系统的安全。

但包过滤技术的缺陷也是很明显的。包过滤技术是一种完全基于网络层的安全技术,只能根据数据包的来源、目标和端口等网络信息进行判断,无法识别基于应用层的恶意入侵,例如恶意的 Java 小程序以及电子邮件中附带的病毒等。有经验的黑客也很容易伪造 IP 地址,骗过包过滤型防火墙。

2. 应用级网关防火墙

应用级网关指的是通常所说的代理服务器。它适用于特定的 Internet 服务,例如超文本传输(HTTP)、远程文件传输(FTP)等。代理服务器通常运行在两个网络之间,阻挡了二者间的数据交流,它对于客户机来说像是一台真的服务器,而对于外界的服务器来说,它又是一台客户机。当客户机需要使用服务器上的数据时,首先将数据请求代理服务器,代理服务器再根据这一请求向服务器索取数据,当代理服务器接收到对某站点的访问请求后会检查该请求是否符合规定,如果规则允许用户访问该站点,代理服务器会像一个客户一样去那个站点取回所需信息再转发给客户。

代理服务器通常都拥有一个高速缓存,这个缓存存储着用户经常访问的站点,在下一个用户要访问同一站点时,服务器就不需要重复地获取相同的内容,直接将缓冲内容发出即可,既节约了时间也节约了网络资源。代理服务器像一堵墙一样挡在内部用户和外界之间,从外部只能看到该代理服务器而无法获知任何的内部资源(例如用户 IP 地址等),外部的恶

意侵害也就很难伤害到企业内部网络系统。应用级网关比单包过滤更为可靠，而且会详细地记录所有的访问状态信息。

但是应用级网关也存在一些不足：它对系统的整体性能有较大影响，使访问速度变慢，因为它不允许用户直接访问网络，而且应用级网关需要对客户机可能产生的每一个特定的 Internet 服务安装相应的代理服务软件，从而大大增加了系统的复杂度；用户不能使用未被代理服务器支持的服务，对每一类服务要使用特殊的客户端软件，但并不是所有的 Internet 应用软件都可以使用代理服务器。

3. 状态监测型防火墙

状态监测型防火墙是新一代的产品，这一技术实际已经超越了最初的防火墙定义。状态监测型防火墙能够对各层的数据进行主动的、实时的监测，在对这些数据加以分析的基础上有效地判断出各层中的非法侵入。这种防火墙具有非常好的安全性，它使用了一个在网关上执行网络安全策略的软件模块，称为检测引擎。检测在不影响网络正常运行的前提下，采用抽取有关数据的方法对网络通信的各层实时监测，抽取状态信息，并动态地保存起来作为以后执行安全策略的参考。检测引擎支持多种协议和应用程序，并可以很容易地实现应用和服务的扩充，同时这种监测型防火墙产品一般还带有分布式探测器，这些探测器安置在各种应用服务器和其他网络的节中，不仅能够检测来自网络外部的攻击，同时对于来自内部网络的恶意破坏也有极强的防御。

与前两种防火墙不同，当用户访问请求到达网关的操作系统前，状态监视器要抽取有关数据进行分析，结合网络配置和安全规定做出接纳、拒绝、身份认证、报警或给该通信加密等处理动作。一旦某个访问违反安全规定，就会拒绝该访问，并报告有关状态作日志记录。状态监测型防火墙的另一个优点是它会检测无连接状态的远程过程调用(RPC)和用户数据报(UDP)之类的端口信息，而包过滤和应用级网关防火墙都不支持此类应用。

这种防火墙无疑是非常坚固的，但它会降低网络的速度，而且配置也比较复杂。好在有关防火墙厂商已注意到这一问题，例如 Checkpoint 公司的防火墙产品 Firewall-1，它所有的安全策略规则都是通过面向对象的图形用户界面(GUI)来定义以简化配置过程。

5.1.3　防火墙的体系结构

1. 屏蔽路由器

屏蔽路由器是一个具有数据包过滤功能的路由器，既可以是一个硬件设备，也可以是一台主机。路由器上安装有 IP 层的包过滤软件，可以进行简单的数据包过滤。因为路由器是受保护网络和外部网络连接的必然通道，所以屏蔽路由器的使用范围很广。但其缺点也非常明显，一旦屏蔽路由器的包过滤功能失效，则受保护网络和外部网络就可以进行任何数据通信了。

2. 双宿主机网关

如果一台主机装有两块网卡，一块连接受保护网络，一块连接外部网络，那么这台堡垒主机就是双宿主机(双重宿主主机)网关，如图 5-2 所示。双宿主机体系结构围绕保垒主机构筑。保垒主机至少有两个网络接口，可以充当与这些接口相连的网络之间的路由器。外部网络能与保垒主机通信，内部网络也能与保垒主机通信，但是外部网络与内部网络不能直

接通信，IP 数据包并不是从一个网络(例如外部网络)直接发送到另一个网络(例如内部网络)，保垒主机的防火墙体系结构禁止这种发送。它们之间的通信必须经过保垒主机的过滤和控制。

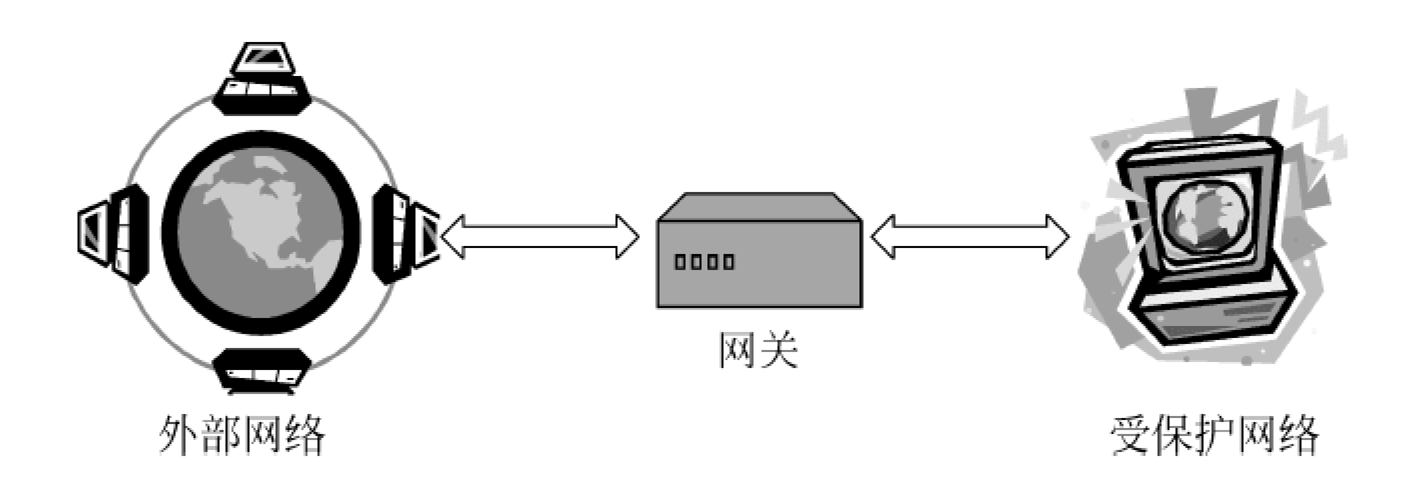


图 5-2 双宿主机网关示意图

堡垒主机装有相应的路由软件，可以很容易地实现网关的功能，并且可以有详尽的日志，也可以安装相应的系统管理软件，便于系统管理员使用。双宿主机网关优于屏蔽路由器的地方是：堡垒主机的系统软件可用于维护系统日志、硬件复制日志或远程日志。这一点对于日后的检查很有用，但不能帮助网络管理者确认内网中哪些主机可能被黑客入侵。双宿主机网关的一个致命弱点是：一旦入侵者侵入堡垒主机并使其只具有路由功能，则任何网络上的用户均可以随便访问内部网络。

3. 被屏蔽主机网关

这种结构由一台屏蔽路由器和一台堡垒主机组成，如图 5-3 所示。

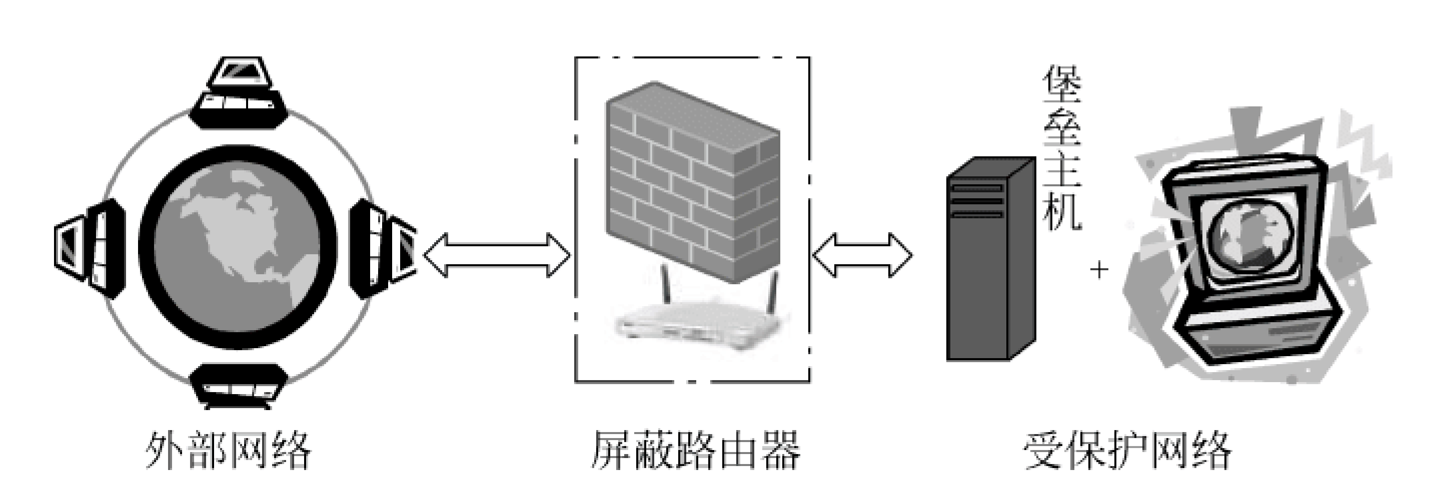


图 5-3 被屏蔽主机网关示意图

堡垒主机在受保护网络中，可以与受保护网络的主机进行通信，也可以和外部网络的主机建立连接。屏蔽路由器的作用是允许堡垒主机和外部网络之间的通信，同时所有受保护网络的其他主机和外部网络直接通信。堡垒主机成为从外部网络唯一可到达的主机，此时它就起到了网关的作用。内部网络的安全由屏蔽路由器和堡垒主机同时保证，如果屏蔽路由器被攻破，则内部网络就直接暴露了。

4. 被屏蔽子网

由两台屏蔽路由器将受保护网络和外部网络隔离开，中间形成一个隔离区(DMZ)，就构成了被屏蔽子网结构，如图 5-4 所示。

隔离区可以被外部网络访问，这一点是由靠近外部网络的屏蔽路由器控制的。企业的 IIS 服务器、FTP 服务器放在隔离区中。外部网络是不能够直接访问内部网络的，这一点由靠近内部网络的屏蔽路由器控制。为了让受保护网络的主机可以和外部网络的主机通信，一般采用的方法是在隔离区内增加一台堡垒主机，这台堡垒主机可以被内部网络的主机访

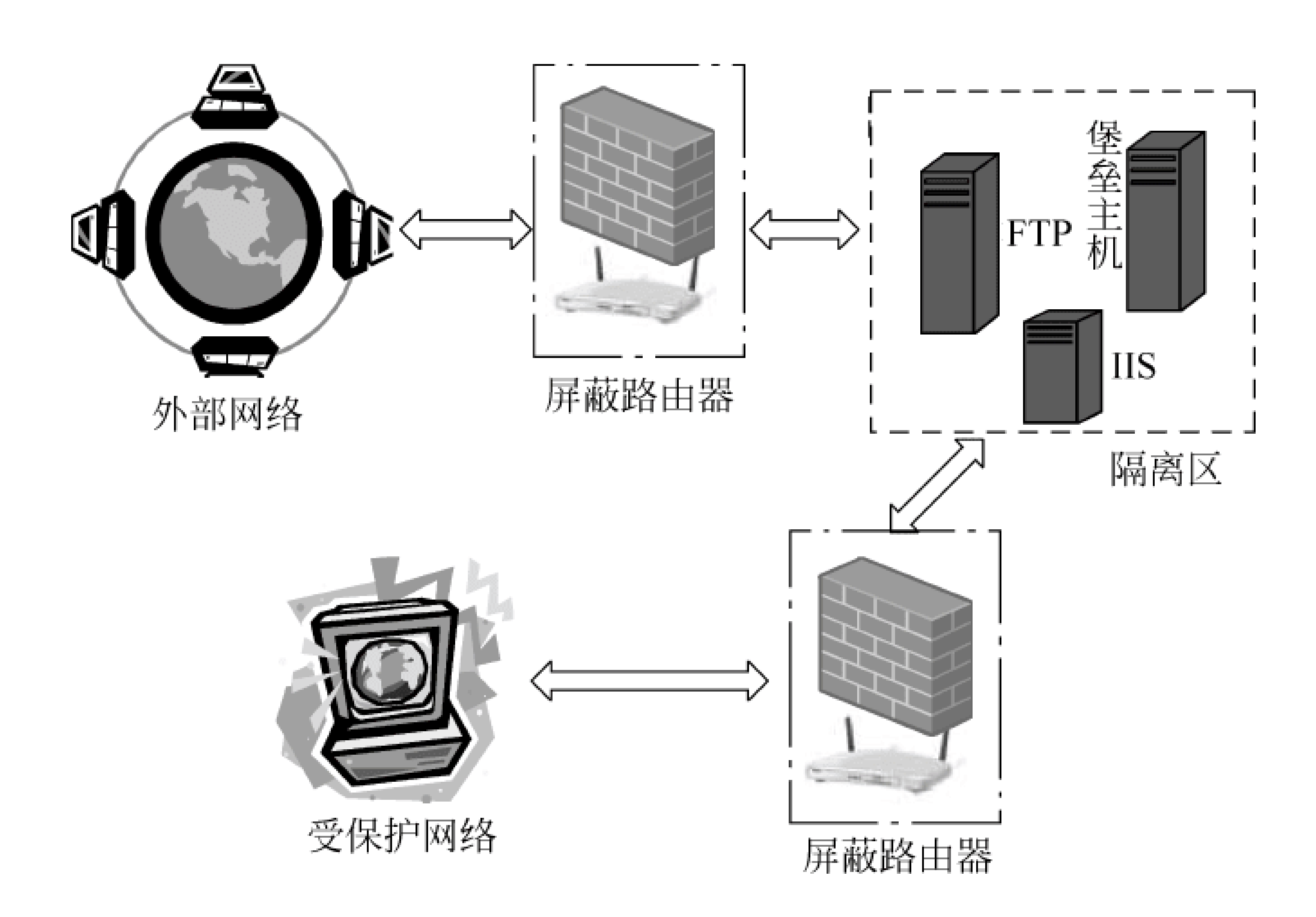


图 5-4　被屏蔽子网结构

问，也可以访问外部网络，此时这台堡垒主机起到了网关的作用，这一点和被屏蔽主机网关的情形类似。这种体系结构比较复杂，但是安全性得到了提升，它将受保护网络的主机和提供服务的服务器隔离起来，使外部网络无法直接到达内部，从而增加了入侵受保护网的难度。

5.1.4　防火墙的功能

随着防火墙技术的不断进步，防火墙的功能也在不断增加，下面介绍几种常见的功能。

1. 包过滤

包过滤功能是防火墙的基本功能，它通过允许或禁止数据包通过防火墙来保证信息安全。对于 5.1.2 节中提出的 3 种类型的防火墙，从本质上来说都是进行了数据包过滤，它们的区别仅仅在于进行包过滤的方法或者位置不同而已。

2. 审计和报警

防火墙具有审计功能是很重要的，安全管理员可能经常要对通过防火墙的信息进行分析，而审计功能的存在就是分析的基础。一般的防火墙会把日志保存到自身或者独立的主机上，后者可以采用更加复杂的分析手段。另外，防火墙也应该具有一定的报警功能，当发现紧急情况时，应该可以通过 E-mail 或手机短信息等方式及时地通知安全管理人员。

3. 代理

代理功能是应用级网关型防火墙的主要功能。一般有两种形式的代理功能：透明代理和传统代理。透明代理可以直接转发受保护网络客户主机的请求，不需要客户主机软件进行相应的设置，对用户保持透明。传统代理则需要客户软件进行必要的设置，最基本的就是要把代理服务器的地址告诉客户软件，5.1.2 节中介绍的应用级网关型防火墙主要就是指的传统代理。

4. NAT

NAT 指的是网络地址转换，主要有两种类型：SNAT（源地址转换）和 DNAT（目的地

址转换)。源地址转换经常用于将保留 IP 地址转换为合法 IP 地址的时候,例如企业内部网络采用保留的 IP 地址,也就是不可路由的 IP 来区分内部主机,当这些主机需要和外部网络进行通信时,就需要转换成一个可以在 Internet 上路由的 IP 地址,这也是源地址转换的典型应用。它既可以解决 IP 地址短缺的问题,又可以对外屏蔽内部网络结构,增加安全性。目的地址转换的一个例子就是刚刚提到的代理功能。

5. VPN

VPN(虚拟专用网络)是近来非常流行的一种功能。随着企业的分布范围越来越广,跨地区的企业网络也越来越多,如果企业的每个部分都采用专线连接,则价格太昂贵,因此大部分企业都采用了 VPN。其实现方法一般是,企业建立 VPN 服务器,外部的办事处或企业分部连接到此服务器上,这条连接一般不采用专线,而是直接通过公共网络,保证传输数据安全的方法是数据加密,IPSec 技术是目前采用的主要技术。

6. 流量统计和控制

防火墙的流量统计功能要求也越来越高,一般的防火墙要实现根据用户的流量统计和根据 IP 地址的流量统计。有了这些统计功能,进行流量控制的要求也就出现了,例如要保证某些 IP 地址的带宽不得低于 10MB 等。

5.1.5 分布式防火墙的实现及应用

1. 分布式防火墙的概念

由于传统防火墙的缺陷不断显露,于是有人认为防火墙是与现代网络的发展不相容的,并认为加密的广泛使用可以废除防火墙。但加密不能解决所有的安全问题,防火墙依然有它的优势,例如通过防火墙可以关闭危险的应用,通过防火墙管理员可以实施统一的监控,也能对新发现的 bug(漏洞)快速做出反应等。也有人提出了对传统防火墙进行改进的方案,例如多重边界防火墙、内部防火墙等,但这些方案都没有从根本上摆脱拓扑依赖,因而也就不能消除传统防火墙的固有缺陷,反而增加了网络安全管理的难度。

个人防火墙的出现弥补了传统防火墙的一些缺陷,它更明确主机会话的上下文关系,同时为网络增加了一道安全屏障,但是它依然无法从根本上解决内部网络的安全问题。原因如下:

(1) 个人防火墙依然依赖网络拓扑结构,容易受 IP 地址欺骗。

(2) 个人防火墙难以统一,网络管理难度大。

(3) 个人防火墙无法实现安全策略的统一配置和管理。

企业中大多数部门员工并非从事计算机行业,为使每个员工掌握防火墙配置技术而对其进行复杂的网络和网络安全知识培训是不现实的。另外,由不精通网络安全知识的员工配置防火墙,导致防火墙形同虚设。因此个人防火墙配合传统防火墙的方案在企业中也同样不可行。

为了克服以上缺陷而又保留防火墙的优点,美国 AT&T 实验室研究员 Steven M Bellovin 教授在他的论文《分布式防火墙》中首次提出了分布式防火墙(Distributed Firewall,DFW)的概念,给出了分布式防火墙的原型框架,奠定了分布式防火墙研究的基础。

传统防火墙缺陷的根源在于它的拓扑结构，分布式防火墙打破了这种拓扑限制，将内部网的概念由物理意义变成逻辑意义。按照 Steven 的说法，分布式防火墙是由一个中心来制定策略，并将策略分发到主机上执行，它使用一种策略语言(例如 Keynote)来制定策略，并被编译成内部形式存于策略数据库中，系统管理软件将策略分发到被保护主机，而主机根据这些安全策略和加密的证书来决定是接收还是丢弃数据包，从而对主机实施保护。在 DFW 中，主机的识别虽然可以根据 IP 地址，但 IP 地址是一种弱的认证方法，容易被欺骗，在分布式防火墙中建议采用强的认证方法，例如 IPSec。加密的证书作为主机认证识别的依据，一个证书的拥有权不易伪造，并独立于拓扑，所以只要拥有合法的证书，不管它处于物理上的内部网还是外部网都被认为是"内部"用户。加密认证是彻底打破拓扑依赖的根本保证。在 DFW 系统中，各台主机的审计事件都被上传到中心日志数据库中统一保存。

2. 分布式防火墙的本质特征

弄清分布式防火墙的本质特征有助于正确认识分布式防火墙，从而划清分布式防火墙和非分布式防火墙之间的界限。

(1) 安全策略必须由管理员统一制定。这是分布式防火墙区别于个人防火墙的根本所在，虽然它们都是主机驻留防火墙，但个人防火墙中的所有行为都是个人行为，别人不能干涉。而分布式防火墙中的行为是集体行为，用户个人不能干涉，每台主机的安全策略都是整个组织安全策略的一部分，全部主机的安全策略之和构成一个组织的整体安全策略，所以分布式防火墙要求实行统一的策略管理。

(2) 策略必须被推到网络的边缘即主机上实施。这是分布式防火墙的又一本质特征，因为分布式防火墙的本意就是要将策略从边界集中实施点迁移到网络末端即主机中来实施。

(3) 日志统一收集管理。因为管理员要对全网进行安全监控，他必须掌握充分的信息，日志是管理员了解信息、追踪攻击者的主要依据。

综上所述，分布式防火墙的本质特征可概括为：策略集中制定分散实施，日志分散产生集中保存。这一本质特征保证了从管理员的角度来看，他管理分布式防火墙就像管理边界防火墙一样，由他负责制定全网的安全策略并对全网的安全状况进行监控，只不过策略的实施不在单一节点上而是分散到了多个节点而已。

3. 分布式防火墙的实现方法

自从 1999 年 11 月 Steven 的《分布式防火墙》发表以来，人们对分布式防火墙的实现进行了研究，提出了一些实现方法，并实现了原型系统，第一个商用分布式防火墙 CyberwallPLUS 也于 2001 年问世。下面介绍分布式防火墙的几种实现方法。

1) 基于 OpenBSD UNIX 的实现

这是提出分布式防火墙概念的 Steven 等人实现的原型系统。该原型系统是在 OpenBSD UNIX 操作系统上修改内核并利用 KeyNote、IPSec 等技术加以实现的。OpenBSD 是理想的开发安全应用的平台。

该原型系统由 3 部分组成：内核扩展模块，用于实施安全机制；用户层策略后台处理程序，用于执行分布式防火墙策略；设备驱动程序，为内核和策略后台程序之间的双向通信提供接口。该原型系统在主机一端的功能模块如图 5-5 所示。

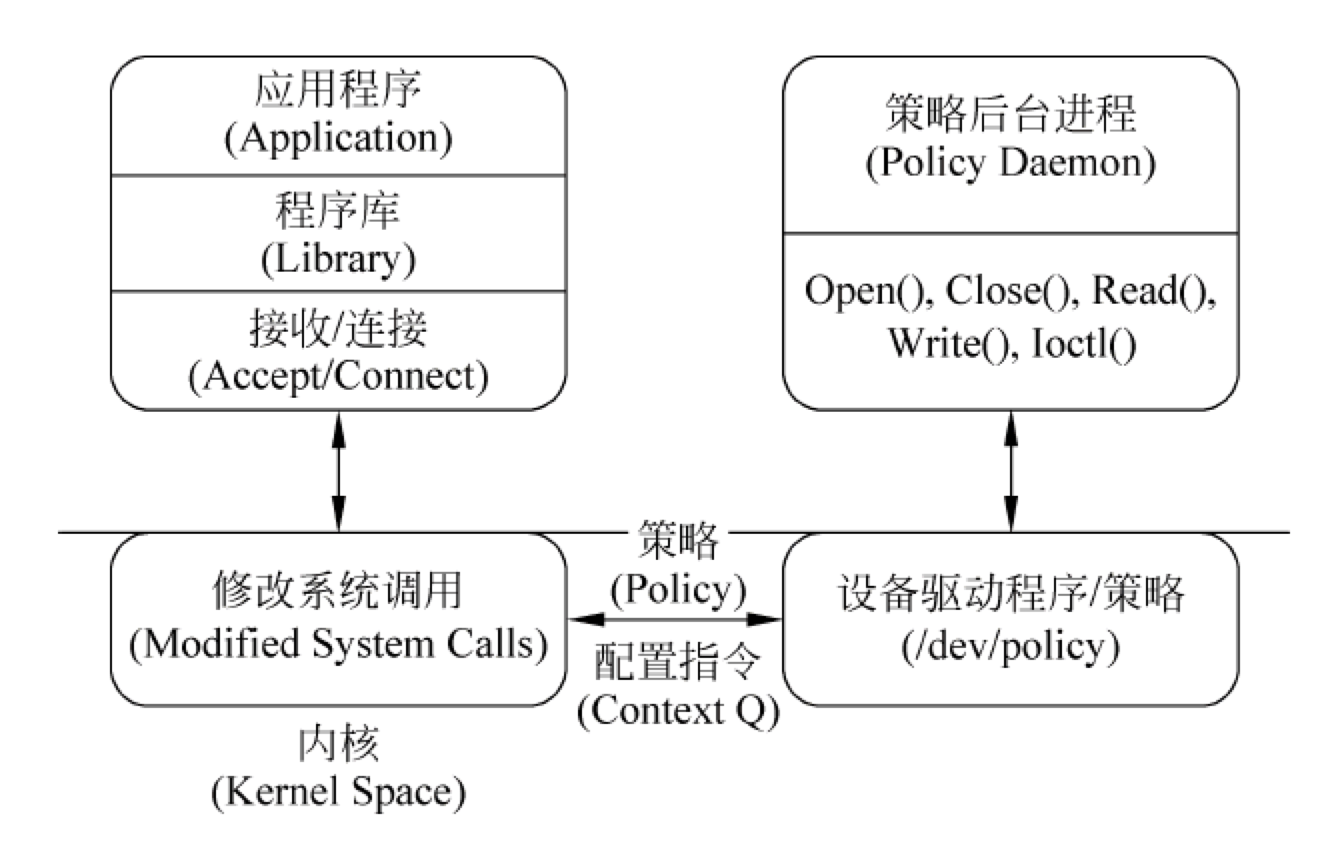


图 5-5 基于 OpenBSD UNIX 的分布式防火墙实现方案

(1) 内核扩展模块：是整个系统的执行模块，其功能是产生并提交策略上下文、根据策略守护进程的答复对数据包进行处理。在 UNIX 系统中用户使用系统调用 connect(2)创建连接请求，使用 accept(2)接收连接请求，一般情况下这两个系统调用不对数据流进行安全检查，为了在内核中实现包过滤功能需要对它进行修改。

(2) 策略后台处理程序：它运行在用户层，作用是根据策略服务器传输过来的安全策略和通信中对方传输的信任书(Credential，相当于证书)来决定接收还是丢弃数据包，并将判断结果返回内核。

(3) 设备驱动程序：该模块的功能是在用户策略后台处理程序和内核中被修改的系统调用之间建立一个通路。它运行于内核态，并向策略后台处理程序提供 read(2)、write(2)等功能调用，后台程序通过调用这些函数与内核交互。

2) 基于 IPsec 的分布式防火墙模型

Steven 在他的《分布式防火墙》中描述了一个基于 IPSec 的分布式防火墙模型。在 Steven 的模型中使用基于 IPSec 的加密证书名称表示网络主机，完全摒弃了以往使用 IP 地址表示主机的方法。该模型共由 3 个部分组成：系统管理模块、翻译器和主机策略执行模块。

网络安全管理员使用系统管理模块来管理所有的主机，定义安全策略，还可以向主机分发新的防火墙软件或安装补丁。网络安全管理员根据主机标识符定义安全策略，然后将定义好的安全策略使用翻译器编译成某种环境的内部格式送出。策略被分发到参与分布式防火墙的各个主机上，有主机策略执行模块负责执行。

3) 基于网卡(NIC)的实现

该方案是美国国防部资助的研究项目，它是基于一种特殊的网卡(3Com3CR990 系列网卡)实现的，称为 EFW(Embedded Firewall，嵌入式防火墙)。这种网卡有内置的处理器和存储器，能独立于主机操作系统而运行；还有内置的加密引擎，使 NIC 之间可以通过 IPSec 加密通信；另外这种网卡使用广泛且价格相对便宜。

(1) EFW 组件：EFW 的主要组件分为主机端组件和服务器端组件，如图 5-6 所示。

① EFW 主机端组件：主要包括 EFW 增强的 NIC、NIC 驱动程序与运行时映像和助理这 3 个部分。EFW 增强的 NIC 中的固件是在安装 EFW 时装入的，它包括包过滤引擎和管

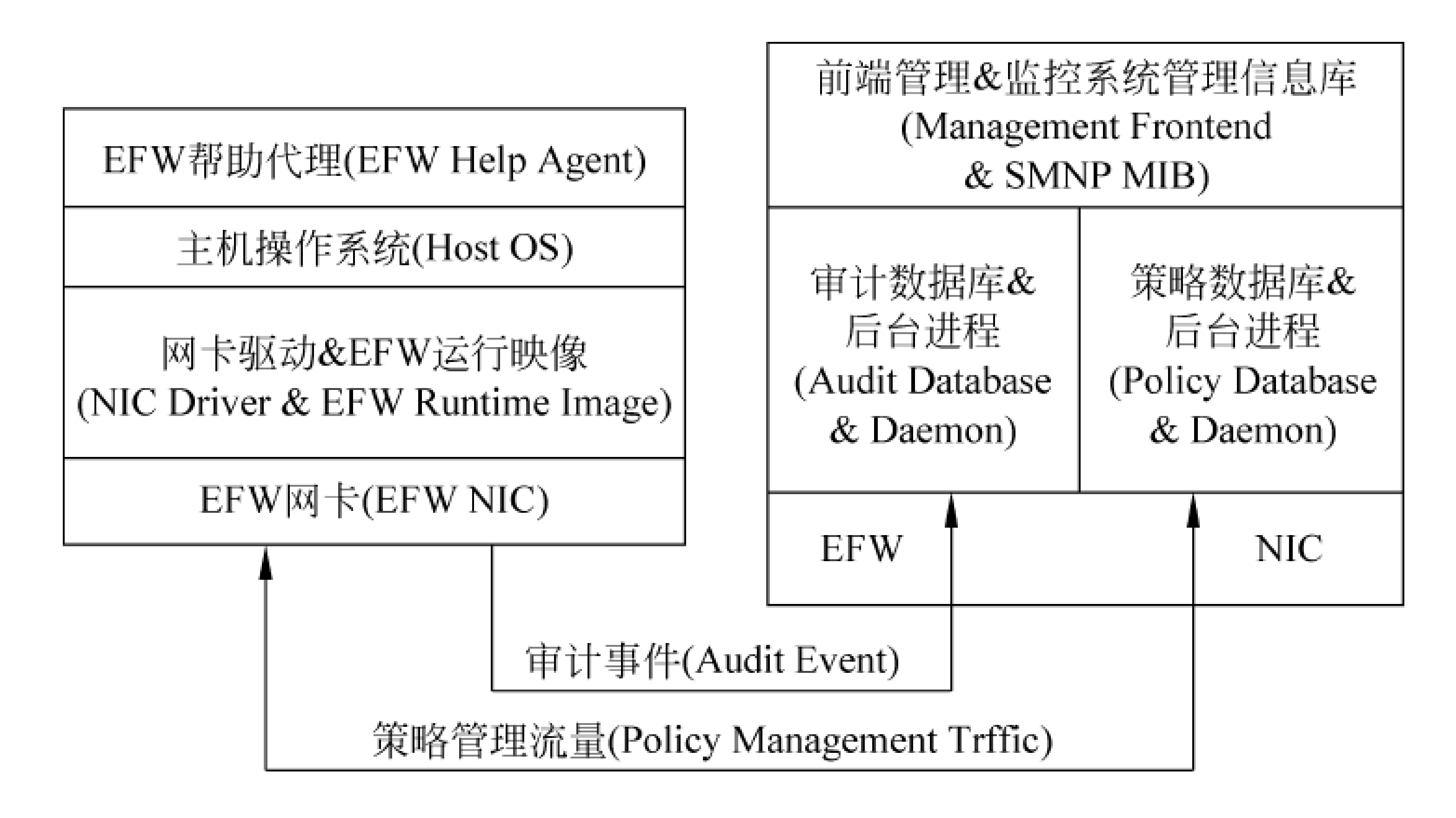


图 5-6　基于网卡的分布式防火墙实现方案

理接口。包过滤引擎能根据标准的参数决定接收或拒绝数据包。管理接口负责从服务器下载策略并上传审计事件到审计数据库，它也负责管理本地 NIC 与服务器之间的安全隧道。驱动程序在机器启动时下载运行时映像（Running Image）并放入固件中，运行时映像的完整性如果受到破坏，NIC 将无法正常工作，从而保证一旦 EFW NIC 被安装，用户将不能废除它，除非通过策略服务器进行适当的操作，因为用户的改动会破坏运行时映像的完整性。EFW 助理的作用有两个，一个作用是给 NIC 传输本机的 IP 地址，另一个作用是定期向策略服务器发送“心跳（Heartbeat）”，这是为了防止恶意的用户用其他 NIC 取代 EFW NIC，因为这样心跳就会停止，从而引起管理员的注意。

② EFW 服务器端组件：主要包括管理组件、策略组件和审计组件 3 个部分。管理组件的主要目的是给管理员提供一个工具去建立和分发策略，也提供一个事件日志浏览器。策略组件的作用是接收管理员定义的策略并编译成过滤规则，然后放入策略数据库中。被保护的机器启动时自动到策略数据库中取回自己的安全策略。当服务器中的策略被修改时，策略组件能自动地将它“推”到相应主机上执行。审计组件收集并整理从各个 NIC 传输过来的审计事件，并提供给管理组件处理。

（2）EFW 的集中管理模式：EFW 将主机分成若干个策略域，每个策略服务器管理一个策略域，一个策略域可以包含整个组织，也可只包含一两个部门。在每个策略域中，NIC 根据主机执行的功能分成若干个组，每个组被分配相同的策略。策略中的规则也可以进行分组，以简化策略的制定、分发和更新。审计事件的设置可建立在整个策略、策略类型或单个策略上。

该实现方案的主要优点是基于硬件、不依赖操作系统，因而难以被绕过，具有坚固的基础。缺陷是 NIC 的处理能力有限。

4）基于 Windows 平台实现的原型系统

CyberwallPLUS 是 Network-1 公司于 2001 年发布的分布式防火墙产品，基于 Windows 平台实现，用于保护 Windows NT/2000 桌面机和服务器。它包括中心管理部件、桌面机防火墙部件和服务器、边界防火墙部件等。所有这些部件都包含如图 5-7 所示的结构，包括包过滤引擎和用户配置接口（可选的）。包过滤引擎采用嵌入内核的方式运行，处于链路层和网络层之间，能够提供访问控制、状态检测和入侵检测的功能。用户配置接口在安

装时是可选的,如果选择安装则用户或管理员可在本地配置安全策略,如果不安装则策略只能由管理员从管理中心加以配置或使用远程管理模块进行配置。该产品实现了中心管理功能,管理员通过中心管理模块可对各台主机实施全方位的控制,包括分发安全策略和远程配置。该产品也具有较完善的审计功能,审计日志可通过建立的连接、阻塞的数据包、入侵尝试和应用类型等来建立(可配置选项)。中心管理模块可对日志和报警信号进行汇集。

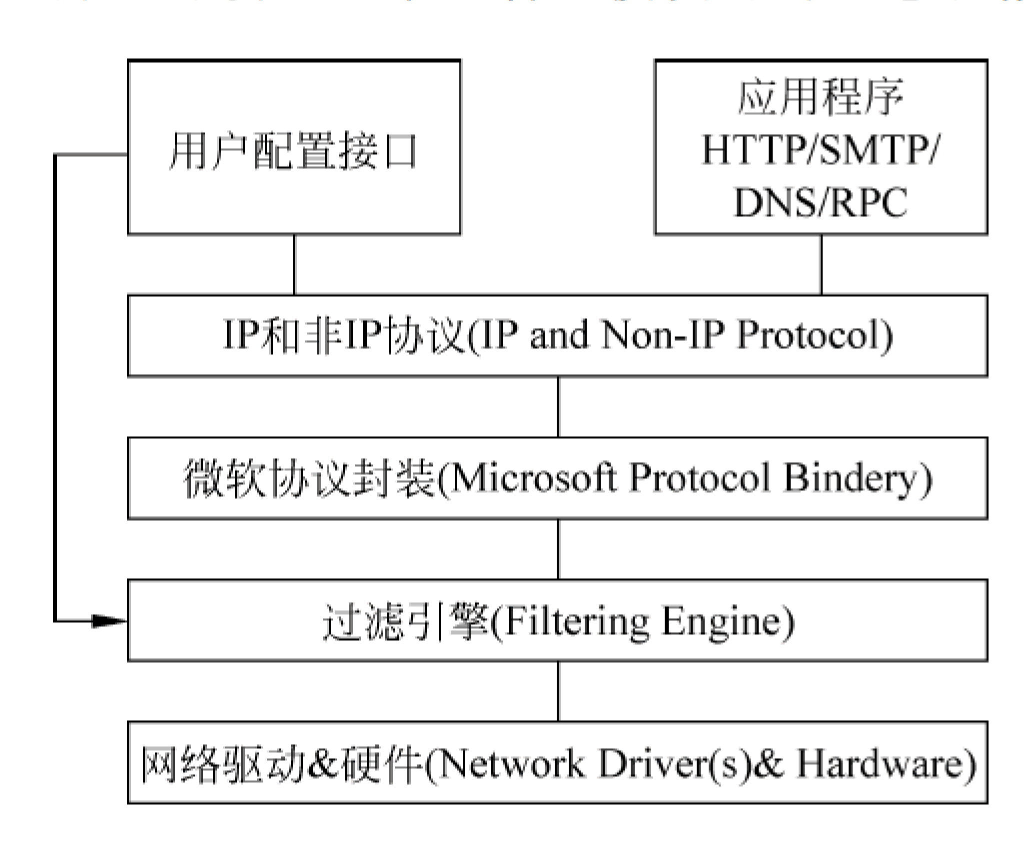


图 5-7 基于 Windows 平台的分布式防火墙实现方案

4. 分布式防火墙的典型应用

对大型网络以及需要打破网络拓扑限制的组织,分布式防火墙是最佳的选择。下面列举两个分布式防火墙的典型应用

1) 锁定关键服务器

对企业中的关键服务器,可以安装分布式防火墙作为第二道防线,使用分布式防火墙的集中管理模块对这些服务器制定精细的访问控制规则,增强这些服务器的安全性。

2) 商务伙伴之间共享服务器

随着电子商务的发展,商务伙伴之间需要共享信息,外联网是一般的解决方案,但外联网的实施代价较高。可以用上面介绍的 EFW 在一台服务器上安装两个 NIC,一个与内部网络相连,另一个与伙伴相连,这样可以方便地实现服务器共享。拥有服务器的一方控制服务器的两个 NIC,分别对其进行设置,使对方能够进入共享服务器,但不能进入本方的内部网络。

5.2 入侵检测技术

5.2.1 入侵和入侵检测

1. 入侵

入侵(Intrusion)是所有试图破坏网络信息的完整性、保密性、可用性、可信任性的行为。入侵是一个广义的概念,不仅包括发起攻击的人取得超出合法范围的系统控制权,也包括收集漏洞信息,造成拒绝服务等危害计算机和网络的行为。入侵行为主要有以下几种:

(1) 外部渗透:指既未被授权使用计算机,又未被授权使用数据或程序资源的渗透。

（2）内部渗透：指虽被授权使用计算机，但是未被授权使用数据或程序资源的渗透。

（3）不法使用：指利用授权使用计算机、数据和程序资源的合法用户身份的渗透。

这 3 种入侵行为是可以相互转变，互为因果的。例如，入侵者通过外部渗透获取了某用户的账户和密码，然后利用该用户的账户进行内部渗透，最后，内部渗透也可能转变为不法使用。

2. 入侵检测

入侵检测（Intrusion Detection）是一种试图通过观察行为、安全日志或审计资料来检测发现针对计算机或网络入侵的技术，这种检测通过手工或专家系统软件对日志或其他网络信息进行分析来完成。而更广义的说法是：识别企图侵入系统非法获得访问权限行为的过程，它通过对计算机系统或计算机网络中的若干关键点收集信息并对其进行分析，从中发现系统或网络中是否有违反安全策略的行为和被攻击的迹象。

入侵检测作为一种积极主动的安全防护技术，提供了对内部攻击、外部攻击和误操作的实时防护，在网络系统受到危害之前拦截和对入侵做出响应。强大的入侵检测软件的出现极大地方便了网络管理，其实时报警功能为网络安全增加了又一道保障。从网络安全立体纵深、多层次防御的角度出发，入侵检测理应受到人们的高度重视，但现状是入侵检测还不够成熟，处于发展阶段，或者是防火墙中集成较为初级的入侵检测模块，所以对于入侵检测技术的研究是很重要的。未来的入侵检测系统将会结合其他网络管理软件，形成入侵检测、网络管理、网络监控三位一体的结构。

5.2.2　入侵检测的分类

1. 特征检测

特征检测又称基于知识的入侵检测，这类检测方法的原则是，任何与已知入侵模型符合的行为都是入侵行为。它要求首先对已知的各种入侵行为建立签名，然后将当前的用户行为和系统状态与数据库中的签名进行匹配。通过收集入侵攻击和系统缺陷的相关知识构成入侵系统中的知识库，然后利用这些知识寻找那些企图利用这些系统缺陷的攻击行为，来识别系统中的入侵行为。系统中任何不能明确地认为是攻击的行为，都可以认为是系统的正常行为。因此，基于入侵知识的入侵检测系统具有很好的检测精确度，至少在理论上具有非常低的虚警率，但是其检测完备性则依赖于入侵攻击和系统缺陷的相关知识的不断更新和补充。

使用这类入侵检测系统，可避免系统以后再遭受同样的入侵攻击，对于网络入侵检测技术的研究可以使系统安全管理员很容易地知道系统遭受到哪种攻击并采用相应的行动。但是，知识库的维护需要对系统中的每一个缺陷都要进行详细分析，这不仅是一个耗时的工作，而且关于攻击的知识，依赖于操作系统、软件版本、硬件平台以及系统中运行的应用程序。这种入侵检测技术主要有以下局限性：

（1）检测系统知识库中的入侵攻击知识与系统运行环境有关。

（2）对于系统内部攻击者的越权行为，由于它们没有利用系统的缺陷，因而很难检测出来。

特征检测的关键问题是规则的获取和表示，构成入侵威胁的审计记录会触发相应规则。

这些规则可以识别出危及系统安全的单个审计事件，也可以分析出构成一个入侵过程的简单审计事件序列。

这种检测方法的特点是检测正确率高而覆盖率偏低，它的弱点是只能发现已知入侵行为。但由于实际情况中大部分入侵者使用的都是已知的攻击方法，因此该技术还是可以有效抵御大部分攻击行为的。

1）专家系统

专家系统是基于知识的检测中运用最多的一种方法。将有关入侵的知识转化成 if-then 结构的规则，即将构成入侵所要求的条件转化为 if 部分，将发现入侵后采取的相应措施转化成 then 部分，当其中某个或部分条件满足时，系统就判断为入侵行为发生。其中的 if-then 结构构成了描述具体攻击的规则库，状态行为以及其语义环境可以根据审计事件得到，推理模块根据规则和行为完成判断工作。

2）状态转换分析

状态转换分析最早由 R. Kemmerer 提出，即将状态转换图应用于入侵行为的分析。状态转换法将入侵过程看作一个行为序列，这个行为序列网络入侵检测技术的研究导致系统从初始状态转入被入侵状态。分析时首先针对每一种入侵方法确定系统的初始状态和被入侵状态，以及导致状态转换的转换条件，和导致系统进入被入侵状态必须执行的操作。然后用状态转换图来表示每一个状态和特征事件，这些事件被集成于模型中，所以检测时不需要一个个地查找审计记录。但是，状态转换是针对事件序列分析的，所以不善于分析过分复杂的事件，而且不能检测与系统状态无关的入侵。

Petri 网用于入侵行为分析是一种类似于状态转换图分析的方法。利用 Petri 网的有利之处在于它能一般化、图形化地表达状态，并且简洁明了。虽然很复杂的入侵特征能用 Petri 网表达得很简单，但是对原始资料匹配时的计算量会很大。下面是这种方法的一个简单示例，如图 5-8 所示，表示在一分钟内如果登录失败的次数超过 4 次，系统便发出警报。其中竖线代表状态转换，如果在状态 S1 发生登录失败，则产生一个标志变量，并存储事件发生时间 T1，同时转入状态 S2。如果在状态 S4 时又有登录失败，而且这时的时间（T2－T1）＜60 秒，则系统转入状态 S5，即为入侵状态，系统发出警报并采取相应措施。

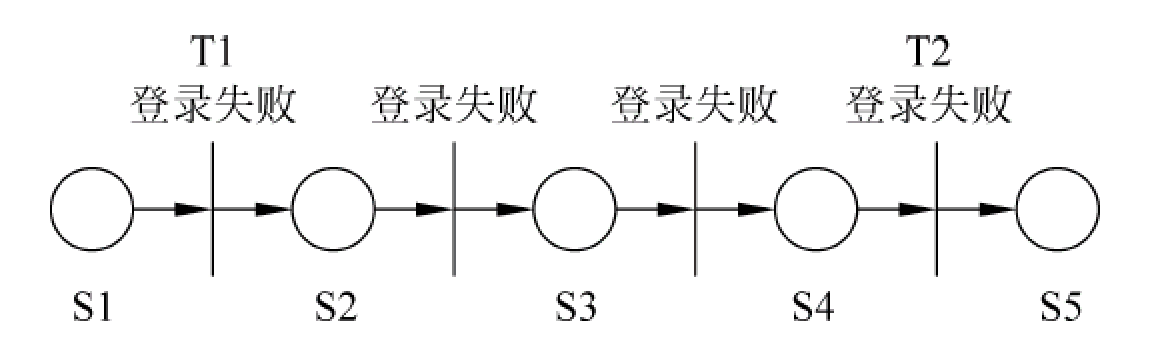


图 5-8　Petri 网一分钟内 4 次登录失败分析

基于知识的检测技术的关键是如何从已知的入侵行为中提取特征，以正确区分真正的入侵与正常行为；如何构造入侵行为的描述语言，并使这种描述语言具有一定的扩展性和适应性。随着对计算机系统弱点和攻击行为的不断收集和研究，入侵行为的特征越来越精确，这使得特征检测技术的使用也越来越广泛。

2. 异常入侵检测

异常检测又称为基于行为的入侵检测，根据使用者的行为或资源使用网络入侵检测技术的研究使用状况来判断是否入侵，而不依赖于是否出现具体行为来检测，任何与已知正常

行为不符合的行为都是入侵行为。这类检测方法的基本思想是：通过对系统审计资料的分析建立起系统主体的正常行为的特征轮廓，检测时，如果系统中的审计资料与已建立的主体正常行为特征有较大出入，就认为系统遭到入侵。特征轮廓是借助主体登录的时间、位置、CPU 的使用时间以及文件的存取属性等，来描述主体的正常行为特征。当主体的行为特征改变时，对应的特征轮廓也相应改变。

异常检测系统在准备阶段通过一定时间的学习为用户正常情况下的行为建立行为轮廓(Profile)，在使用阶段系统一方面通过比较用户当前行为与原先行为轮廓的偏差来检测入侵，另一方面继续根据用户的正常行为来修正行为轮廓。同样，异常检测系统也可为整个计算机系统建立正常行为轮廓。

异常入侵检测方法的关键在于对用户或者系统建立正确的行为轮廓，在早期的异常入侵检测系统中通常用统计模型来进行，例如将用户登录时间、登录失败次数、资源访问频度等一些特征量作为随机变量，通过统计模型计算出这些随机变量的新观察值落在一定区间内的概率，并且根据经验规定一个阈值，超过阈值则认为发生了入侵。后来有很多人工智能技术应用于异常检测，例如神经元网络技术和资料挖掘技术等。

异常入侵检测的最大优点是能检测出一些未知攻击，最大缺点是会产生很大的虚警率，因为异常并不一定是入侵，而且结果缺乏可解释性。

1）概率统计方法

概率统计方法是基于异常检测中应用最早也是最多的一种方法。检测器根据用户对象的动作为每个用户都建立一个用户特征表，通过比较当前特征与已存储的固定模式的以前特征，从而判断是否是异常行为。

用户特征表需要根据审计记录情况不断地加以更新。用于描述特征的变量类型有：

(1) 操作密度：度量操作执行的速率，常用于检测通过长时间平均觉察不到的异常行为。

(2) 审计记录分布：度量在最新记录中所有操作类型的分布。

(3) 范畴尺寸：度量在一定动作范畴内特定操作的分布情况。

(4) 数值尺度：度量那些产生数值结果的操作，例如 CPU 使用量、I/O 使用量等。

这些变量所记录的具体操作包括：CPU 的使用，I/O 的使用，使用地点及时间，邮件使用，编辑器使用，编译器使用，所创建、删除、访问或改变的目录及文件，网络上的活动等。在 SRI/CSL 的入侵检测专家系统中给出了一个特征简表的结构：

<变量名，行为描述，例外情况，资源使用，时间周期，变量类型，门限值，主体，客体，值>

其中的变量名、主体和客体唯一地确定了每一个特征简表，特征值由系统根据审计资料周期性地产生。这个特征值是所有有悖于用户特征的异常程度值的函数。如果假设 S_1，S_2，…，S_n 分别是用于描述特征的变量 M_1，M_2，…，M_n 的异常程度值，S_i 值越大说明异常程度越大。则这个特征值可以用所有 S_i 值的加权平方和来表示：

$$M = \sum_{i=1}^{n} a_i S_i^2, \quad a_i > 0$$

其中，a_i 表示每一个特征的权值。

如果选用标准偏差作为判别准则，则标准偏差为：$\sigma^2 = M/(n-1) - \mu^2$，其中 $\mu = M/n$。如果某 S 值超过了 $\mu \pm d\sigma$，就认为出现了异常。

概率统计方法的优越性在于能够发现未知的入侵,并能对用户活动进行适应性学习,以发现内部用户的渗透和异常,有成熟的概率统计理论基础。但也有些不足之处:统计检测对事件发生的次序不敏感,完全依靠统计理论可能漏检那些利用彼此关联事件的入侵行为;定义是否入侵的判断阀值的选择困难,如果该值设的过高则漏检率提高,如果阀值过低则会造成误检率提高。

2)神经网络方法

利用神经网络检测入侵的基本思想是用一系列信息单元(命令)训练神经单元,这样在给定一组输入后,就可能预测出输出。与统计理论相比,神经网络更好地表达了变量间的非线性关系,并且能自动学习和更新。实验表明 UNIX 系统管理员的行为几乎全是可以预测的,对于一般用户,不可预测的行为也只占了很少的一部分。用于检测的神经网络模块结构大致是这样的:当前命令和刚过去的 w 个命令组成了网络的输入,其中 w 是神经网络预测下一个命令时所包含的过去命令集的大小。根据用户的代表网络入侵检测技术的研究性命令序列训练网络后,该网络就形成了相应用户的特征表,于是网络对下一事件的预测错误率在一定程度上反映了用户行为的异常程度。基于神经网络的检测思想可用图 5-9 表示。

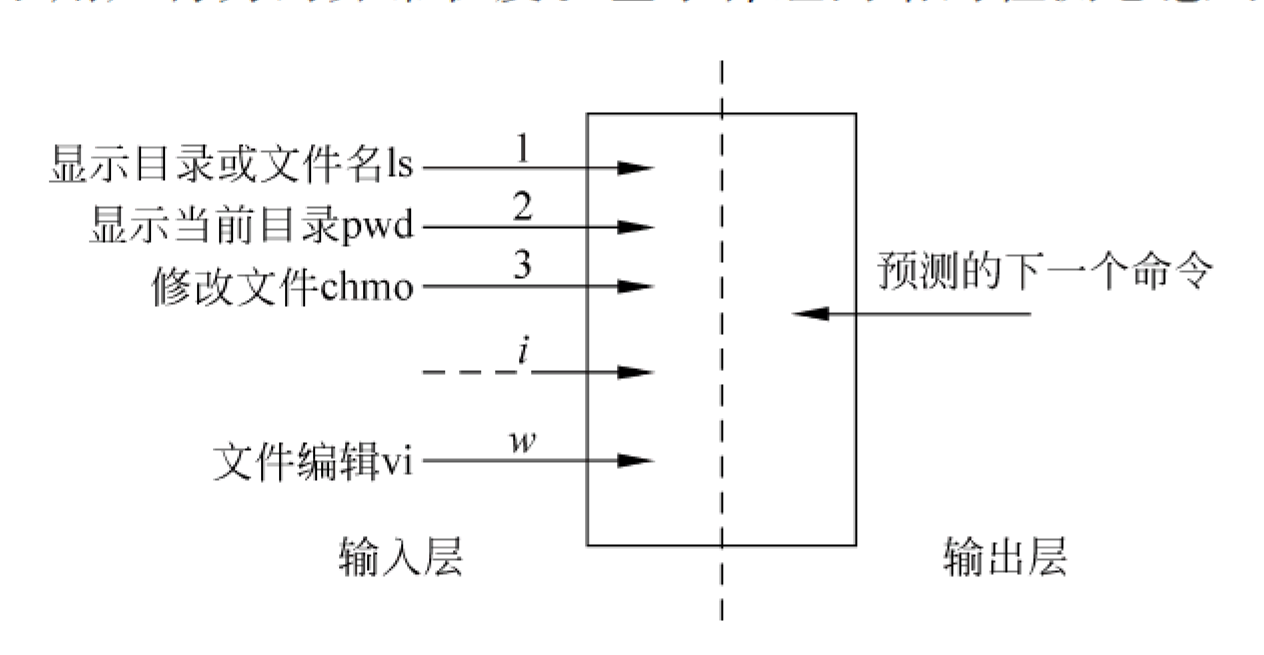


图 5-9 神经网络的检测思想

图 5-9 中输入层的 w 个箭头代表了用户最近的 w 个命令,输出层预测用户将要发生的下一个动作。神经网络方法的优点在于能更好地处理原始资料的随机特性,即无须对这些资料作任何统计假设,并且有较好的抗干扰能力。缺点在于网络拓扑结构以及各元素的重复性很难定;命令窗口 w 的大小也难以选取,窗口太小则网络输出不好,窗口太大则网络会因为大量无关资料而降低效率。

如果能在一个检测系统中将异常入侵检测和特征的入侵检测有机地结合起来,那么会大大减少入侵检测的虚警率和漏警率,基于安全规范的入侵检测方法可以起到这种作用,它的优点是不仅能识别已知攻击,还能识别出未知的攻击。

5.2.3 入侵检测系统及其分类

最早的入侵检测模型由 Dorothy Denning 在 1968 年提出。这个模型与具体输入无关,对此后的大部分实用系统都很有借鉴价值。入侵检测系统(Intrusion Detection System,IDS)在逻辑上必须包含最基本的 3 个部分:数据提取模块、数据分析模块和结果处理模块。入侵检测一般分为 3 个步骤:数据提取、数据分析和结果处理。入侵检测系统基本结构如图 5-10 所示。

图 5-10 中模块划分是基于功能的划分,省略了界面管理模块、配置管理模块等其他

模块。

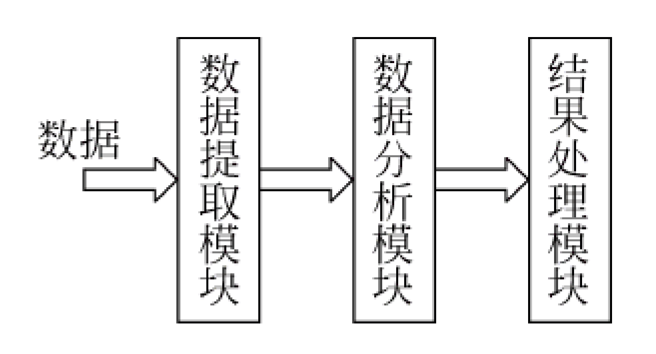


图 5-10　入侵检测系统的基本结构图

数据提取模块的作用在于为系统提取数据。数据为网络数据包、计算机的日志文件和系统调用记录等。如果系统是基于主机的入侵检测系统,数据主要为主机的日志文件、审计记录等。如果系统是基于网络的入侵检测系统,数据就为网络中传输的数据包。

数据分析模块对数据进行深入分析,发现攻击并根据分析的结果产生事件,传递到结果处理模块。数据分析的方式有很多种,并大致分为误用检测和异常检测两大类型。数据分析模块是入侵检测系统的核心模块。

结果处理模块的作用在于告警与反应,这实际上与 PPDR 模型的 R 有所重叠。结果处理模块应该对不同的攻击有不同的响应策略,一般发现攻击后,模块会启动一些相对应的事件,例如通知管理员、系统自动恢复以前的状态、切断网络等。

入侵检测系统的具体原理如图 5-11 所示。

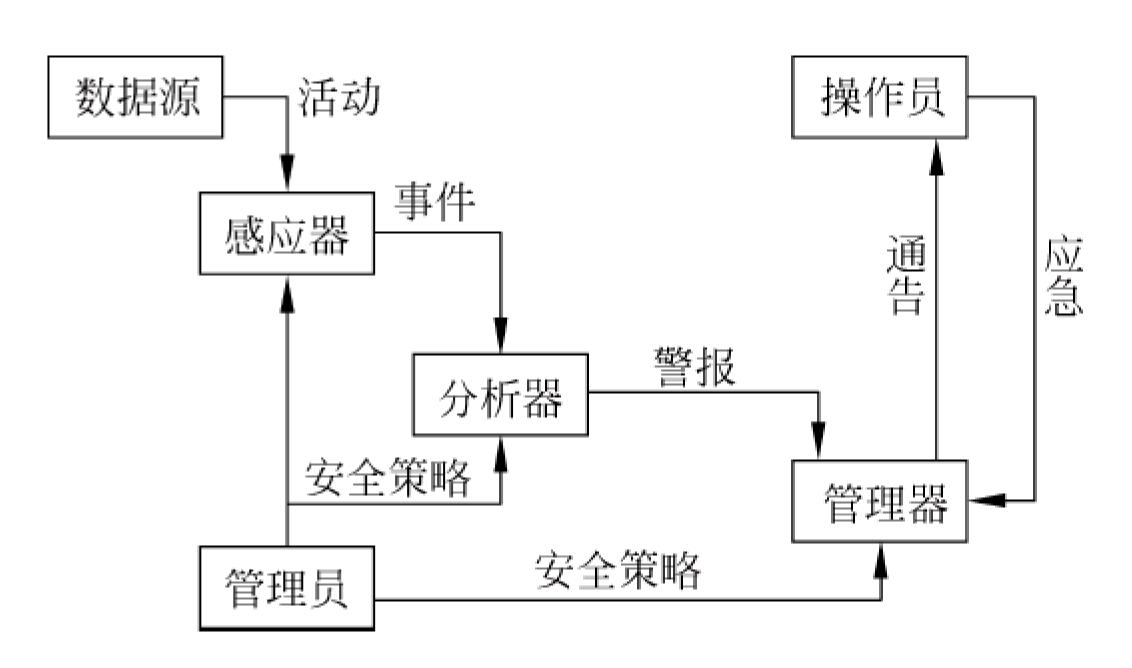


图 5-11　入侵检测系统原理示意图

入侵检测系统的原理比较简单,当感应器感知到数据后,由系统管理员所提供的安全策略对该感应事件进行分析审计数据,一旦分析结果认定为入侵则发出警报信息,启动管理器通知操作人员并启动相应的应急措施,如关闭相应连接、切断网络,以便帮助管理员采取进一步的应急措施。

目前国内外已经开发出许多入侵检测系统,这些 IDS 从不同的角度有着不同的分类方法,其中最主要的是按照入侵检测的信息来源和检测方法来进行分类的。

1. 基于主机和网络的入侵检测系统

入侵检测系统根据信息来源的不同可以分为基于主机的入侵检测系统(Host-based Intrusion Detection System,HIDS)和基于网络的入侵检测系统(Network-based Intrusion Detection System,NIDS)两大类。基于主机的入侵检测系统从单个主机上提取资料(例如系统日志等)作为入侵分析的资料源,而基于网络的入侵检测系统从网络上提取网络报文作为入侵分析的资料源。通常来说基于主机的入侵检测系统只能检测单个主机系统,而基于网络的入侵检测系统可以对本网段的多个主机系统进行检测,多个分布于不同网段上的基于网络的入侵检测系统可以协同工作以提供更强的入侵检测能力。

1) 基于主机的入侵检测系统(HIDS)

基于主机的入侵检测系统(HIDS)主要从主机的审计记录和日志文件中获得所需的数

据,并辅以主机上的其他信息,例如文件系统属性、进程管理状态、系统资源使用情况等,在此基础上完成检测入侵行为的任务。基于主机的入侵检测在20世纪80年代初期就出现了,早期的入侵检测系统都是基于主机的入侵检测技术。那时网络环境比较简单,在这种情况下通过记录检查可疑行为是非常常见的操作。由于入侵在当时是相当少见的,在对攻击的事后分析就可以防止今后的攻击。HIDS在发展过程中还结合了一些其他技术,对关键的系统文件和可执行文件的检查是入侵检测的一个常用方法,主要是通过定期检查校验和来进行,以便发现意外的变化;还有一些监测端口活动的方法,通过监测特定端口,当发现它们被访问时向管理员报警。

HIDS的主要目的是在事件发生后提供足够的分析研究来阻止进一步的攻击,尽管它不如NIDS快捷,但它确实具有NIDS无法比拟的优点。这些优点包括:

(1) 能够监视特定的系统行为。HIDS能够监视所有的用户登录和退出,甚至用户所做的所有操作,审计系统在日志里记录的策略改变,监视关键系统文件和可执行文件的改变等。

(2) 因为检测在主机上运行的命令序列要比检测网络流相对简单得多,系统的复杂性也小得多。HIDS通常情况下比NIDS的虚警率要低,由于使用含有已发生事件信息,HIDS可以比NIDS更加准确地判断攻击是否成功。

(3) 有些攻击在网络的数据流中很难发现,或者根本没有通过网络而是在本地进行,这时NIDS将无能为力,只能借助于HIDS。

(4) 适用被加密的和交换的环境。由于HIDS安装在遍布企业的各种主机上,它们比NIDS更加适于交换和加密的环境。交换设备可将大型网络分成许多小型网络段加以管理,所以从覆盖足够大的网络范围的角度出发,很难确定配置NIDS的最佳位置。HIDS可安装在所需的重要主机上,在交换的环境中具有更高的能见度。某些加密方式也向NIDS发出了挑战。根据加密方式在协议堆栈中的位置不同,NIDS可能对某些攻击没有反应,而HIDS没有这方面的限制,当操作系统及HIDS发现即将到来的业务时,数据流已经被解密了。

HIDS的主要缺点有:

(1) HIDS安装在需要保护的设备上,这会降低应用系统的效率。因为它依赖于服务器固有的日志与监视能力,如果服务器没有配置日志功能则必须重新配置,否则会给运行中的业务系统带来不可预见的性能影响。

(2) 全面布置HIDS代价较大。企业中很难将所有主机都采用HIDS保护,只能选择部分主机保护。那些尚未安装HIDS的机器将成为保护的盲点,入侵者可以利用这些机器达到攻击目标。

(3) HIDS除了监测自身的主机以外,根本不检测网络上的情况。而且对入侵行为的分析的工作量会随着主机数目的增加而增加。

2) 基于网络的入侵检测系统(NIDS)

基于网络的入侵检测系统(NIDS)通常是作为一个独立的单元放置于被检测网络上的。它使用原始网络数据包作为入侵检测的数据来源,通常利用一个运行在混杂模式下的网络适配器来实时监视并分析网络中所有通信数据。NIDS通常使用4种常用检测技术来识别入侵:模式、表达式或字节匹配;频率或阈值判断;低级事件的相关性;统计学意义上的非常规现象检测。一旦检测到了攻击行为,NIDS的响应模块就提供多种选项以通知、报警并对攻击采取相应的反应。反应会因产品而异,但通常都包括通知管理员、报警、中断连接和

作为证据支持起诉而做的会话记录。

NIDS 的主要优点有：

(1) 成本较低。NIDS 可在几个关键访问点上进行策略配置，以观察发往多个系统的网络通信。

(2) 能够检测到 HIDS 无法检测的入侵。NIDS 能够检查数据包的头部而发现非法攻击；能够检测到那些来自网络的攻击；能够检测到超过授权的非法访问。

(3) NIDS 不依赖于保护主机的操作系统，而且隐蔽性好。一个网络上的监测器不像一个主机那样显眼和易被存取，因而也不那么容易遭受攻击。

NIDS 的主要缺点有：

(1) 只检查它直接连接网段的通信，不能检测在不同网段的网段包。在使用交换以太网的环境中就会出现检测范围的局限，而安装多台 NIDS 的传感器会使部署整个系统的成本大大增加。

(2) NIDS 可能会将大量的数据传回分析系统中。在一些系统中监测特定的数据包会产生大量的分析数据流量。一些系统在实现时采用一定方法来减少回传的资料量，对入侵判断的决策由传感器实现，而中央控制台成为状态显示与通信中心，不再作为入侵行为分析器。这样的系统中的传感器协同工作能力较弱。

(3) 网络入侵检测系统处理加密的会话过程较困难。目前通过加密信道的攻击尚不多，随着 IPv6 的普及，这个问题会越来越突出。

这两种入侵检测系统都具有自己的优点和不足，互相可作为补充。HIDS 可以精确地判断入侵事件，可对入侵事件立即进行反应，还可以针对不同操作系统的特点判断应用层的入侵事件，其缺点是会占用主机宝贵的资源。而 NIDS 只能监视本网段的活动，并且精确度较差，在交换网络环境中难于配置，防入侵欺骗的能力也比较差，但是它可以提供实时网络监视，并且监视粒度更细致。

随着高速网络和交换式网络的迅速发展，入侵检测领域的实践者们倾向于将这两种不同的方法结合起来，IDS 必须包含紧密融合的主机和网络部分，以便得到更多的攻击和入侵信息。必须大幅度提高网络对攻击和错误使用的抵抗力，使安全措施的实施更加有效，并使设置更加灵活。目前出现了一种网络节点入侵检测系统(Network Node Intrusion Detection System，NNIDS)，它结合了上述两种方法，将入侵检测任务委派到网络中的各个主机上，以减轻由于高速和交换式网络而给入侵检测系统带来的巨大压力，同时它还非常适用于网络中存在加密资料的情况。

2. 误用和异常入侵检测系统

入侵检测系统根据检测方法可分为两种基本检测类型：误用入侵检测系统(Misuse Intrusion Detection System)和异常入侵检测系统(Anomaly Intrusion Detection System)。

1) 误用入侵检测系统

误用入侵检测系统根据已知入侵类型(知识、模式等)来检测目标网络系统中的入侵，它是指运用已知攻击方法，根据已定义好的入侵模式，通过分析数据判断这些入侵模式是否出现来进行检测。首先收集入侵行为的特征，建立相关的误用模式库，在后续的检测过程中，将收集到的数据与库中的特征代码进行比较，得出是否入侵的结论。这种方法由于依据具体特征库进行判断，所以检测准确度很高，并且因为检测结果有明确的参照，也为系统管理

员做出相应措施提供了方便。

误用检测主要不足在于只能检测已知的攻击模式，当新漏洞或新入侵方式出现时，需要由人工或其他机器学习系统得出新入侵行为的特征模式，添加到误用模式库中，才能使系统具备检测新的入侵行为的能力。

2）异常入侵检测系统

异常入侵检测系统将被监控系统正常行为的信息作为检测目标系统中是否有入侵的异常活动的依据，它根据使用者的行为或资源使用状况的正常程度来判断是否入侵。其特点是首先总结正常操作应该具有的特征，得出正常操作的模型，然后对后续的操作进行监视，一旦发现偏离正常统计等意义上的行为，立即进行报警。异常检测的优点是它能抽取系统的正常行为以此检测系统异常行为。这种能力不受系统以前是否知道这种入侵与否的限制，所以能够检测新的入侵行为。

异常入侵检测的主要不足则是误报率很高。此外，若入侵者了解到检测方法，就可以通过慢慢训练检测系统，避免系统指标突变，到最后连异常行为也可能认为是正常的方法来进行欺骗以达到入侵目的。

误用入侵检测系统通常需要定义一组规则，而这种工作模式不能发现新的攻击行为，故不能提供全面的保护，但误报率很低。异常入侵检测系统所能检测到的威胁行为更多，包括已知的和未知的威胁，但这种模式会导致大量的误报。通过这一比较，不难发现它们在很大程度上具有互补性。要在提高检测率的同时避免过高的误报率，就必须将两种检测方法有效地结合起来。

3. 集中式和分布式入侵检测系统

根据入侵检测系统各模块运行的分布方式不同，可分为集中式入侵检测和分布式入侵检测两类。

1）集中式入侵检测系统（Centralized Intrusion Detection System，CIDS）

CIDS的各个模块包括数据的收集与分析以及响应都集中在一台主机上运行，这种方式适用于网络环境比较简单的情况。CIDS也可以有多个分布于不同主机上的审计程序，但只有一个中央入侵检测服务器，审计程序将当地收集到的数据踪迹发送给中央服务器进行分析处理。CIDS在系统的可伸缩性、可配置性方面存在致命缺陷。随着网络规模的增大，主机审计程序和服务器之间传输的数据量就会骤增，必将导致网络性能的降低。而且一旦中央服务器出现故障，整个系统将会陷入瘫痪。

2）分布式入侵检测系统（Distributed Intrusion Detection System，DIDS）

相对于CIDS，DIDS的各个模块分布在网络中不同的计算机、设备上，其分布性主要体现在数据收集模块上，如果网络环境比较复杂、数据量比较大，那么数据分析模块也会分布在网络的不同计算机和设备上，通常是按照层次性的原则进行组织。DIDS根据各组件间的关系还可细分为层次式DIDS和协作式DIDS。其中层次式DIDS是一种部分分布控制形式，而协作式DIDS是全分布式控制形式。

（1）层次式DIDS：为解决CIDS的缺陷，在监视大规模网络时，需将网络进行分层管理。在层次式DIDS中，定义了若干个分等级的监测区域，每一个区域有一个专门负责分析数据的IDS，每一级IDS只负责所监测区域的数据分析，然后将结果传输给上一级IDS。层次式DIDS通过分层分析很好地解决了集中式IDS的不可扩展的问题，但同时也存在下列

问题：当网络的拓扑结构改变时，区域分析结果的汇总机制也需要做相应的调整；一旦位于最高层的 IDS 受到攻击后，其他那些从网络多路发起的协同攻击就容易逃过检测，造成漏检。

（2）协作式 DIDS：协作式 DIDS 将中央检测服务器的任务分配给若干个互相合作的 HIDS，这些 HIDS 不分等级，各司其职，负责监控本地主机的某些活动，所有的 HIDS 并发执行并相互协作。协作式 DIDS 的特点就在于它的各个节点都是平等的，一个局部 DIDS 的失效不会导致整个系统的瘫痪，也不会导致协同攻击检测的失败。因而，系统的可扩展性、安全性都得到了显著提高。但同时它的维护成本也很高，并且增加了所监控主机的工作负荷，例如通信机制、审计开销、踪迹分析等。而且主机之间的通信、审计以及审计数据分析机制的优劣直接影响了协作式 DIDS 的效率。

5.2.4　入侵检测系统的局限性及发展趋势

1. 当前的入侵检测产品存在的问题

虽然入侵检测系统的重大作用不言而喻，但是它作为一项比较新的技术，还存在一些技术上的困难，不是所有厂商都有研发入侵检测产品的实力。目前的入侵检测产品大多存在这样一些问题。

1）误报和漏报的矛盾

入侵检测系统对网络上所有的数据进行分析，如果攻击者对系统进行攻击尝试，而系统相应服务开放，只是漏洞已经修补，那么这一次攻击是否需要报警？这就是一个需要管理员判断的问题，因为这也代表了一种攻击的企图。但大量的报警事件会分散管理员的精力，反而无法对真正的攻击做出正确反映。和误报相对应的是漏报，随着攻击的方法不断更新，入侵检测系统是否能报出网络中所有的攻击也是一个问题。

2）隐私和安全的矛盾

入侵检测系统可以收到网络的所有数据，同时可以对其进行分析和记录，这对网络安全极其重要，但难免会对用户的隐私构成一定风险，这就要看具体的入侵检测产品是否能提供相应功能以供管理员进行取舍。

3）被动分析与主动发现的矛盾

入侵检测系统采取被动监听的方式发现网络问题，无法主动发现网络中的安全隐患和故障。如何解决这个问题也是入侵检测产品面临的问题。

4）海量信息与分析代价的矛盾

随着网路数据流量的不断增加，能否高效处理网路中的数据也是衡量入侵检测产品的重要依据。

5）功能性和管理性的矛盾

随着入侵检测产品功能的增加，可否在功能增加的同时尽可能地降低管理难度？例如，入侵检测系统的所有信息都存储在数据库中，此数据库能否自动维护和备份而不需管理员的干预？另外，入侵检测系统自身安全性如何？是否易于部署？采用何种报警方式？这些都是需要考虑的因素。

6）单一的产品与复杂的网络应用的矛盾

入侵检测产品最初的目的是为了检测网络的攻击，但仅仅检测网络中的攻击远远无法

满足目前复杂的网络应用需求。通常，管理员难以分清网络问题是由于攻击引起的还是网络故障引起的。入侵检测系统检测出的攻击事件又如何处理？可否和目前网络中的其他安全产品进行配合？

2. 入侵检测系统的发展趋势

未来 DIDS 技术的发展将着重于以下几个方面。

1）分析技术的改进

入侵检测误报和漏报的解决将最终依靠分析技术的改进。目前入侵检测分析方法主要有统计分析、模式匹配、数据重组、协议分析和行为分析等。

统计分析是统计网络中相关事件发生的次数，达到判别攻击的目的。模式匹配利用对攻击的特征字符进行匹配完成对攻击的检测。数据重组是对网络连接的数据流进行重组再加以分析，而不仅仅分析单个数据包。

协议分析技术是在对网络数据流进行重组的基础上，理解应用协议，再利用模式匹配和统计分析技术来判明攻击。例如某个基于 HTTP 协议的攻击含有 ABC 特征，如果此数据分散在若干个数据包中，假如一个数据包包含 A，另外一个包含 B，还有一个包含 C，则单纯的模式匹配就无法检测，只有基于数据流重组才能完整检测。而利用协议分析则只在符合的协议（例如 HTTP）检测到此事件才会报警。假设此特征出现在 E-mail 里，因为不符合协议，就不会报警。利用此技术有效地降低了误报和漏报。

行为分析技术不仅简单分析单次攻击事件，还根据前后发生的事件确认是否确有攻击发生，攻击行为是否生效，这是入侵检测分析技术的最高境界。但目前由于算法处理和规则制定的难度很大，目前还不是非常成熟，但这是入侵检测技术未来发展的趋势。目前最好综合使用多种检测技术，而不只是依靠传统的统计分析和模式匹配技术。另外，规则库是否及时更新也和检测的准确程度相关。

2）内容恢复和网络审计功能的引入

前面已经提到，入侵检测的最高境界是行为分析。但行为分析目前还不是很成熟，因此个别优秀的入侵检测产品引入了内容恢复和网络审计功能。内容恢复即在协议分析的基础上，对网络中发生的行为加以完整的重组和记录，网络中发生的任何行为都逃不过它的监视。网络审计即对网络中所有的连接事件进行记录。入侵检测的接入方式决定入侵检测系统中的网络审计不仅类似防火墙可以记录网络进出信息，还可以记录网络内部连接状况，此功能对内容无法恢复的加密连接尤其有用。

内容恢复和网络审计让网络管理员看到网络的真正运行状况，其实就是调动网络管理员参与行为分析的过程。此功能不仅能使网络管理员看到孤立的攻击事件的报警，还可以看到整个攻击过程，了解攻击确实发生与否，查看攻击者的操作过程，了解攻击造成的危害，不但发现已知攻击，同时发现未知攻击，不但发现外部攻击者的攻击，也发现内部用户的恶意行为。毕竟网络管理员是最了解其网络的，他们通过此功能的使用，很好地达到了行为分析的目的。但使用此功能的同时需注意对用户隐私的保护。

3）集成网络分析和管理功能

入侵检测不仅可以对网络攻击进行检测，同时还可以收到网络中的所有数据，对网络的故障分析和健康管理也可起到重大作用。当网络管理员发现某台主机有问题时，也希望能马上对其进行管理。入侵检测也不应只采用被动分析方法，最好能和主动分析相结合。所以，

入侵检测产品集成网管功能、扫描器(Scanner)及嗅探器(Sniffer)等功能是以后发展的方向。

4）安全性和易用性的提高

入侵检测系统是个安全产品，其自身安全极为重要。因此，目前的入侵检测产品大多采用硬件结构透明式接入来免除自身安全问题。同时，对易用性的要求也日益增强，例如，全中文的图形界面，自动的数据库维护，多样的报表输出。这些都是优秀入侵检测产品的特性和以后继续发展细化的趋势。

5）改进对大数据量的网络的处理方法

随着对大数据量处理的要求，入侵检测产品的性能要求也逐步提高，出现了千兆入侵检测等产品。但如果入侵检测产品不仅具备攻击分析，同时具备内容恢复和网络审计功能，则其存储系统也很难完全工作在千兆环境下。这种情况下，网络数据分流也是一个很好的解决方案，性价比也较好。这也是国际上较通用的一种作法。

6）防火墙联动功能

入侵检测系统发现攻击，自动发送给防火墙，防火墙加载动态规则拦截入侵，称为防火墙联动功能。目前此功能还没有到完全实用的阶段，主要是一种概念，随便使用会导致很多问题。目前主要的应用对象是自动传播的攻击，例如 Nimda 等，联动只在这种场合有一定的作用。无限制地使用联动，若未经充分测试，对防火墙的稳定性和网络应用会造成负面影响。但随着入侵检测产品检测准确度的提高，联动功能日益趋向实用化。

思 考 题

(1) 计算机网络有哪些漏洞？

(2) 什么样的网络是安全的？网络安全的重要性有哪些？

(3) 简述网络安全所涉及的主要技术。

(4) 简述防火墙在网络安全中的地位，它可以分为几种类型？

(5) 典型的防火墙有哪些方面的基本特征？

(6) 防火墙有哪些不足？

(7) 入侵检测技术弥补了防火墙的哪些不足？

(8) 试描述通用的入侵检测系统的基本结构。

(9) 简述基于主机的入侵检测系统的特点。

(10) 简述基于网络的入侵检测系统的优缺点。

(11) 根据检测原理入侵检测系统可以分为几种？它们的原理分别是什么？

(12) 简述入侵检测技术的发展方向。

参 考 文 献

[1] 袁艺，张晓燕，卫红. 网络战在平时悄然打响. 保密工作，2010，(10)：52～54.

[2] 柯科峰，邵世煌. 企业入侵检测系统的研究与实现. 计算机应用研究，2004，(1)：154～158.

[3] 周秋霞，梁启文. 校园网络入侵检测系统的设计与实现. 重庆工学院学报(自然科学版)，2007，(12)：58～60.
[4] 宿洁，袁军鹏. 防火墙技术及其进展. 计算机工程与应用，2004，(9)：147.
[5] 刘学波，孟丽荣. 高速网络环境下的入侵检测系统的研究. 计算机工程与设计，2005，(5)：6～38.
[6] 潘永刚. 浅谈入侵检测系统的应用与趋势. 鄂州大学学报，2008，(52)：25～27.
[7] 孙雷. 入侵检测系统在计算机网络安全上的应用. 应用能源技术，2009，(8)：45～46.
[8] 张国华，肖频. 一种基于网络的入侵检测系统设计. 微计算机信息，2009，(2)：70～72.
[9] 王文奇，郑秋生，吴婷. 高速入侵检测研究. 计算机工程与设计，2008，(14)：3616～3620.
[10] 宋劲松. 网络入侵检测——分析、发现报告攻击. 北京：国防工业出版社，2004.
[11] 杨琼，杨建华，王习平，马斌. 基于防火墙与入侵检测互动技术的系统设计. 武汉理工大学学报，2005，(7)：113～115.
[12] 曹天杰. 计算机系统安全. 北京：高等教育出版社，2003.
[13] 朱林平，万郡. 浅谈入侵检测系统. 计算机与现代化，2006，(12)：121～123.
[14] Pacek TH，Newsham TN. Insertion，Evasion，and Denial of Service：Eluding Network Intrusion Detection. Se2cureNetworks，Inc，January，1998.
[15] 丁志芳，徐孟春，汪淼，殷石仓. 关于入侵检测系统与入侵防御系统的探讨. 光盘技术，2006，(1)：21～23.
[16] Stephen Northcutt. 网络入侵检测分析员手册. 北京：人民邮电出版社，2000.
[17] 张丽红，赵俊忠. 计算机网络入侵检测系统发展趋势. 计算机测量与控制，2004，(4)：301～305.
[18] 杜彦辉. 利用蜜罐技术实现对互联网非法活动进行监控. 中国人民公安大学 47 参考文献学报(自然科学版)，2007，(4)：48～50.
[19] 樊雷. 校园网入侵检测系统的研究和设计. 福建电脑，2008，5：117～118.
[20] Julia Allen，Alan Christie，William Fithen. State of the Practice of Intrusion Detection Technologies. Networked Systems Survivability Program，2000.
[21] John Chrirllo. Hack Attacks Revealed. America Copyright China Machine Press，2003.
[22] 祝晓光. 网络安全设备与技术. 北京：清华大学出版社，2004.
[23] 汪静，王能. 入侵检测系统设计方案的改进. 计算机应用研究，2004，(7)：208～210.
[24] 王永波，梅波. 浅谈计算机网络入侵检测系统. 黑龙江科技信息，2008，(12)：95～97.

第 6 章　数据备份与恢复技术

本章学习目标

在全球信息化飞速发展的今天,数据已经成为一项非常重要的资产,对数据的保护日益重要。本章将介绍数据的基本概念及数据备份与恢复所需的相关基础知识,并对当前数据保护技术进行比较全面的介绍,具体包括数据备份技术与数据恢复技术。

通过对本章的学习,应掌握以下内容:

(1) 数据备份与数据恢复的必要性和原则。

(2) 数据存储方式和数据结构方式。

(3) 数据备份策略和技术。

(4) 硬盘的基础知识和硬盘数据恢复技术。

随着计算机技术的发展,计算机系统被广泛地应用于日常事务,众多的个人、公司、企业甚至国家机关的数据都被保存在计算机中,例如各类管理中所产生的机密办公文件、经营中所积累的客户资料以及研发中所涉及的重要文件等。如果由于某些原因导致这些数据丢失,将会给用户造成严重的损失。因此保护数据存储的一致性、完整性及可用性,对用户来说是至关重要的。

本章将从数据的基本概念入手,介绍数据备份和数据恢复的相关定义及所需的基础知识,并对常见的数据安全技术(数据备份技术和数据恢复技术)进行全面介绍。

6.1　数据备份与恢复概述

本章中所说的数据仅指计算机中的数据。在计算机系统中,数据是指各种字母、数字符号的组合,包括用户数据和系统数据。用户数据主要是指计算机文件系统中的各种软件或文档资料以及文字、声音和图像等组成的多媒体文件等。系统数据主要是指各种计算机系统数据,例如计算机系统中各种硬件及系统信息等。数据存储在计算机系统中,要保证数据的一致性、完整性及可用性,才能让数据为用户所用。保证数据安全可用的主要方法是数据备份(Backup)技术与数据恢复(Recovery)技术。

数据备份就是保存数据的副本,以期望达到预防事故(例如自然灾害、病毒破坏和人为损坏等)造成的数据损失的目的。数据恢复就是将数据恢复到事故之前的状态。数据备份与数据恢复总是相对应的,备份是恢复的前提,恢复是备份的目的,而无法恢复的备份是没有任何意义的。下面将对数据备份和数据恢复技术的概念、特性和方法进行阐述。

6.1.1　数据安全的主要威胁

数据库中的数据丢失或被破坏的原因主要有以下几点:

(1) 计算机硬件设备故障。由于硬件设备质量或硬件老化等原因,计算机硬件可能会

出现故障，导致不能使用。例如硬盘损坏会使得存储其中的数据丢失。

(2) 软件故障。由于软件设计上的失误或用户使用的不当，软件系统可能会误操作数据引起数据破坏。

(3) 人为误操作。例如用户不小心使用了诸如 delete、update 等命令而引起数据丢失或被破坏。

(4) 数据窃取。由于 Internet 上黑客的入侵或者来自内部网的蓄意破坏。

(5) 病毒入侵。破坏性病毒会破坏系统软件、硬件和数据。

(6) 自然灾害。例如火灾、洪水、地震或雷击等自然灾害会对计算机系统及其数据造成极大的破坏。

6.1.2 数据备份概述

尽管数据备份已经出现了相当长时间，但依然没有统一的定义。虽然对数据备份的定义有多种说法，但是其核心思想是相同的。这里给出了数据备份一个较为公认的定义。

数据备份是指将全部或部分数据集合从应用主机的硬盘或阵列复制到其他存储介质的过程，目的是为了在设备发生故障或发生其他威胁数据安全的灾害时保护数据，以便重新加以利用，将数据遭受破坏的程度减到最小。例如，在日常生活中，人们常会将认为重要的文件多保存几份，这就是一个简单的备份过程。但在实际的企业应用中，数据备份绝不仅仅是简单的文件复制，将会涉及系统、数据库等多种信息，是一个相对较复杂的过程。

1. 数据备份的重要性

数据备份是由存储介质、操作系统、备份软件和其他备份工具为保护数据建立起的一项安全措施。给数据备份，其实就是相当于给数据买了一份保险。在信息时代，企业业务的发展离不开信息系统，特别是银行、证券、保险等行业，其信息系统中的数据就是企业的核心价值。有关研究表明，丢失 300MB 的数据对于市场营销部门就意味着 13 万元人民币的损失，对财务部门就意味着 16 万元的损失，对工程部门来说损失可达 80 万元。而丢失的关键数据如果 15 天内仍得不到恢复，企业就有可能被淘汰出局。

然而，尽管数据非常重要，但很多计算机信息系统并没有增加保护重要数据的预算，也几乎没有采取必要措施以保证发生灾难后能继续开展业务。据调查显示，目前国内企业中只有不到 15%的服务器连有备份设备，而国际上，以美国为首的发达国家都非常重视数据存储备份技术，企业中服务器与磁带机的连接超过 60%，许多企业把数据备份放在对信息系统投入预算的第一位。

相比国外对数据备份技术的重视，可以看出国内用户对数据备份的认识不深，对保护数据的了解也不多，对整个数据存储管理和备份缺乏专业和系统的考虑，这是非常危险的。因此增强数据备份的意识，加强数据备份操作对企业是非常必要的。

2. 数据备份的原则

从数据备份的定义中可以看出数据备份的最终目的是当数据发生灾难时，可以恢复数据，而不能恢复数据的备份系统是没有意义的。因此在制定备份系统时，要注意能够将数据恢复过来。在实际应用中，有多种计算机系统应用环境，不同的应用环境则要求不同的解决方案来适应，一般来说，一个完善的备份系统需要满足以下原则。

1）稳定性

备份系统的主要作用是为系统提供一个数据保护的方法，因此该备份系统本身的稳定性和可靠性就变成了一个重要方面。一方面，备份系统一定要能与操作系统 100%兼容，若不能与操作系统兼容，其备份功能就无从谈起了；另一方面，当灾难事故发生时，能够快速有效地恢复数据，若不能快速地恢复数据，在这个竞争激烈的时代，用户无法承担长时间等待的代价。

2）全面性

在复杂的计算机系统环境中，可能会包括各种操作平台，例如 UNIX、NetWare、Windows NT、VMS 等，并安装了各种应用系统，例如 ERP、数据库和群件系统等，备份系统应该能维持各种操作系统、数据库及其典型应用。只有备份系统具有了这样的全面性，才能满足用户的多种需求。

3）自动化

很多系统由于工作性质，对何时备份、用多长时间备份都有一定的限制。在下班时，系统负荷轻，适于备份，可是这会增加系统管理员的负担，由于其精神状态等原因，还会给备份安全带来潜在的隐患。因此，备份方案应能提供定时的自动备份，并利用磁带库等技术进行自动换带。在自动备份过程中，还要有日志记录功能，并在出现异常情况时自动报警。

4）高性能

随着业务的不断发展，数据越来越多，更新越来越快，在休息时间来不及备份如此多的内容，在工作时间备份又会影响系统性能。这就要求在设计备份时，尽量考虑到提高数据备份的速度，例如利用多个磁带机并行操作等备份方法。

5）维持业务系统的有效性

实时备份对业务系统的性能将会产生一定的影响，有时会很大。如何采取有效的技术手段避免备份对服务器系统、数据库系统、网络系统的影响，将是非常重要的。

6）操作简单

数据备份应用于不同领域，进行数据备份的操作人员也处于不同的层次。这就需要一个直观的、操作简单的图形化用户界面，缩短操作人员的学习时间，减轻操作人员的工作压力，使备份工作得以轻松地设置和完成。

7）实时性

有些关键性的任务是要 24 小时不停地运行的，在备份的时候，有一些文件可能仍然处于打开状态。那么在进行备份的时候，要采取措施，实时地查看文件大小、进行实物跟踪，以保证正确地备份系统中的所有文件。

8）容灾考虑

将磁带库中的磁带复制一份，存放在远离数据中心的地方，以防止数据中心发生不可预测的灾难。

3. 数据备份的要求

遵循上述备份系统的原则，对数据备份有下列要求：

(1) 数据备份实现自动化，以减少系统管理员的工作量。

(2) 数据备份工作应制度化、科学化。

(3) 介质管理的有效化，防止读写操作的失误。

（4）数据分门别类保存到存储介质中，使数据的存储更细致、科学。

（5）自动介质的清晰轮转，提高介质的安全性和使用寿命。

（6）对各种平台的应用系统以及其他信息数据进行集中备份，系统管理员可以在任意一台工作站上管理、监控、配置备份系统，实现分布处理、集中管理。

（7）维护人员可以容易地恢复被破坏的整个文件系统和各类数据。

（8）备份系统还应考虑网络带宽对备份性能的影响，备份系统平台的选择及安全性、备份系统容量的适度冗余，备份系统良好的扩展性等因素。

6.1.3 数据恢复概述

类似对数据备份定义的情况，长期以来，计算机领域对数据恢复也没有一个标准定义。但在多种定义里，其基本思想是一致的。这里也给出了一个对数据恢复较为公认的概念。

数据恢复，是指当存储介质出现损伤或由于人员误操作、操作系统本身故障所造成的数据看不见、无法读取、丢失时，将损坏的数据还原成正常数据，恢复它本来的“面目”的过程；数据恢复不仅对已丢失的文件进行恢复，还可以恢复物理损伤的磁盘数据，也可以恢复不同操作系统的数据。

数据备份是保护数据安全的一种预防措施，而数据恢复就是出现问题之后的一种补救措施。数据备份是为数据恢复服务的，其终极目标就是数据恢复。但是由于数据出现的问题有多种情况，数据备份策略并不能满足所有问题，所以并不是只要做了数据备份，就能轻松地将所需的数据恢复过来，并且针对没有备份的情况，数据恢复就更困难了。因此，对数据恢复技术的研究也引起了广泛关注。

在现实中，由于操作人员或计算机系统的不同，导致数据出现的问题千差万别，所以在进行数据恢复时，一定要谨慎细致，对每一步的操作都有一个明确的目的。要考虑该操作是否可行，是否必要，是否具有破坏性。对于有可能破坏数据的操作，要特别谨慎，在操作之前最好进行备份，以防万一。

在开始恢复数据之前，应该首先完成以下几个步骤：

（1）备份当前尚能工作的驱动器上的所有数据。若 C 盘损坏，那么在开始任何工作之前首先备份 D 盘（及其他盘）上的数据到其他可靠的地方。

（2）将损坏的硬盘拿到一个正常工作的相同的操作系统下，如果条件不允许，取下该硬盘，并安装一个新的主硬盘，在重新挂上损坏硬盘之前对硬盘分区并格式化，立即更改 COMS 设置。

（3）调查使用者。查出在丢失数据之前发生的事情，查出是否有其他应用程序对磁盘进行过操作。最后的用户输入非常重要，要查出使用者在送交磁盘前做过些什么。

（4）如果可能，备份所有扇区是一个非常不错的方法。按文件进行的转存在这里没有任何帮助。如果进行克隆，确保是按位进行而不是按文件进行。

（5）手头要有一个好的扇区编辑工具，例如 WinHex 就是一款不错的基于扇区的编辑工具。

（6）尽可能多地得到最后使用者的关键文件信息。

完成这些步骤后，就会对数据的损坏情况有一个基本的了解，例如会了解出现这个问题的原因、数据的破坏程度有多少。根据这些前提知识，就可以制定初步的恢复计划，例如怎

样进行恢复才能达到最好的恢复效果,主要步骤有哪些等。

6.2　数据备份

6.2.1　数据备份模式

根据备份操作的层次,可以将数据备份模式分为文件级备份和块级备份。

文件级备份也称为逻辑备份,只能感知到文件这一层,将磁盘上所有的数据以文件的形式读出,备份到另一个介质上。这些文件在原来的介质上存放可以是不连续的,各个不连续的块之间的链关系由文件系统来管理。而将这些文件备份到存储介质上后,该文件的备份数据的存放就是连续的。恢复数据的时候,软件会重构磁盘文件系统,并从存储介质读出数据,向磁盘写入数据。

块级备份也称为物理备份,需要备份块设备上的每个块,不管这个块上有没有数据,或者这个块上的数据属于哪个文件。块级备份忽略了文件的结构,处理过程较简洁,直接对磁盘扇区进行读取,每次备份数据都是以一个扇区(512B)为单位来进行备份,原设备有多少容量,就备份多少容量,并将读取到的扇区写入用于备份的存储介质。

两种备份模式由于其基本机制不同各有不同特点,在实际应用中,要根据需求来决定使用何种备份模式。两种备份模式的特点比较如下。

1. 备份速度

块级备份是面向物理存储设备的,不需要经过操作系统的文件系统接口,而是通过磁盘控制器驱动接口直接读取磁盘,所以相对文件级的备份来说速度有所加快。

2. 磁盘开销

因为块级备份会备份许多空扇区,所以块级备份所备份的数据量相对文件级备份要多。另外,文件级备份会将原来不连续的文件备份成连续存放的文件,恢复的时候也会在原来的磁盘上连续写入,所以很少造成碎片。而块级备份在备份之后,原来不连续的文件在备份系统的存储介质上的存放还是不连续的,恢复的时候也只是将块的状态原样恢复,碎片数量不会减少。

3. 操作效率

文件级备份在对非连续存储磁盘上的文件进行备份时需要额外的查找操作。这些额外的操作增加了磁盘的开销,降低了磁盘的吞吐率。所以,跟块级备份相比较,文件级备份性能较差。

文件级备份模式下,文件即使一个很小的改变也需将整个文件备份,对于一个文件很大的情况下,就会大幅度地降低备份效率,增加磁盘开销和备份时间。而块级备份避免了当文件出现一个小的改动时就需要对整个文件做备份,只是会去做改动部分的备份,有效地提高了备份效率,节省了备份时间。

4. 实时性

块级备份可以做到高效的实时备份,因为在每次主机往磁盘写入数据的时候,都需要同时将数据写入到备机,这种写入操作都是基于磁盘扇区的,所以很快就能被识别。只有在写

入到备机完成之后，才会返回给上层的应用系统来继续下一步工作。

文件级备份是很难做到实时备份的，因为它的每次修改都是基于文件的，而文件的哪部分被修改，系统很难实时捕获到，所以备份的时候需要把整个文件读一遍再发到备机，实时效率不高。

5. 支持度

块级备份是在文件系统之下对数据进行复制，所以它不受文件系统限制，可以支持各种文件系统包括未格式化(RAW)分区。文件级备份是以单个文件为单位对数据进行复制，所以它受文件系统限制，仅能对部分支持的文件系统做备份，不支持 RAW 分区。

6.2.2 数据存储方式

数据的存储方式针对不同的应用环境有不同的解决方案，主要可分为直接附加存储(Direct Attached Storage，DAS)、存储区域网(Storage Area Networking，SAN)和网络附加存储(Network Attached Storage，NAS)3 种。

1. 直接附加存储(DAS)

DAS 也可称为服务器附加存储(Server-Attached Storage，SAS)，是直接连接于主机服务器的一种存储方式。每一台主机服务器有独立的存储设备，每台主机服务器的存储设备无法互通，需要跨主机存取资料时必须经过相对复杂的设定，若主机服务器分属不同的操作系统，要存取彼此的资料则更加复杂，有些系统甚至不能存取。DAS 通常用在单一、数据交换量不大且性能要求不高的网络环境下，如图 6-1 所示。

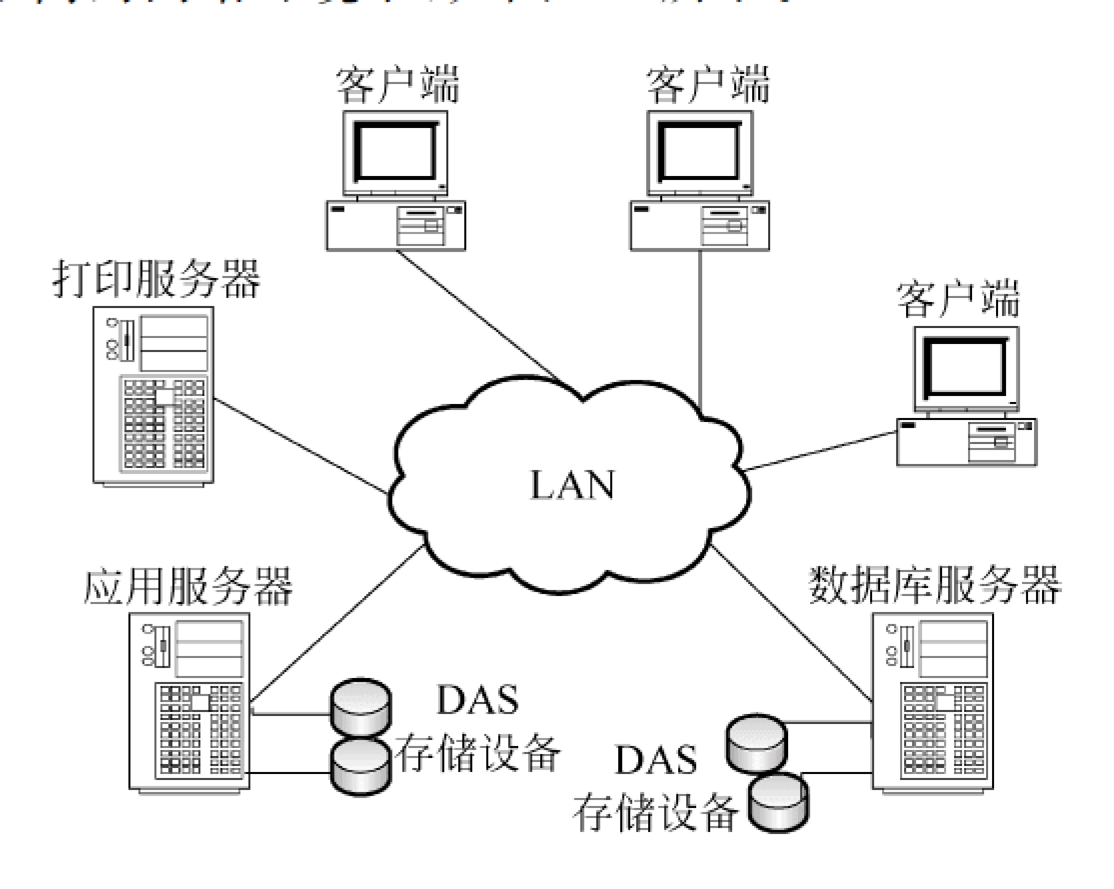


图 6-1 基于 DAS 结构的存储备份

2. 存储区域网(SAN)

SAN 是一种用高速(光纤)网络连接专业主机服务器的一种存储方式。此系统会位于主机群的后端，它使用高速 I/O 连接方式，例如 SCSI、ESCON 及 Fiber-Channels 等，提供 SAN 内部任意节点之间的多路可选择的数据交换，如图 6-2 所示。一般而言，SAN 应用在对网络速度要求高、对数据的可靠性和安全性要求高、对数据共享的性能要求高的应用环境中，特点是代价高，性能好。例如，电信、银行的大数据量存储应用。

3. 网络附加存储(NAS)

NAS 又称为网络直连存储,它是一种专业的网络文件存储及文件备份设备,通常是直接连接在网络上并提供资料存取服务。一套 NAS 存储设备就如同一个提供数据文件服务的系统,特点是性价比高。例如,教育、政府、企业等数据存储应用。基于 NAS 结构的存储备份如图 6-3 所示。

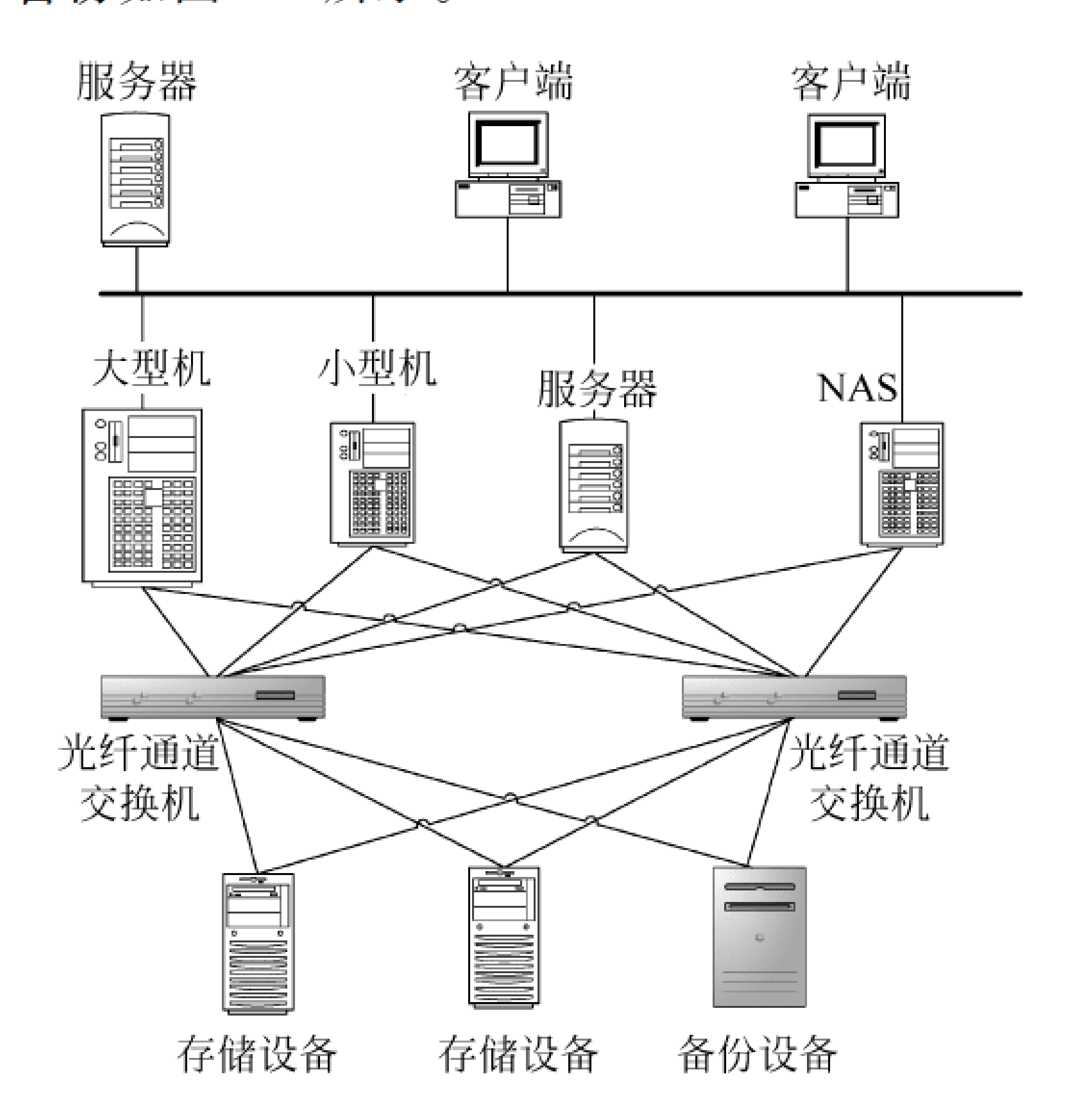


图 6-2　基于 SAN 结构的存储备份

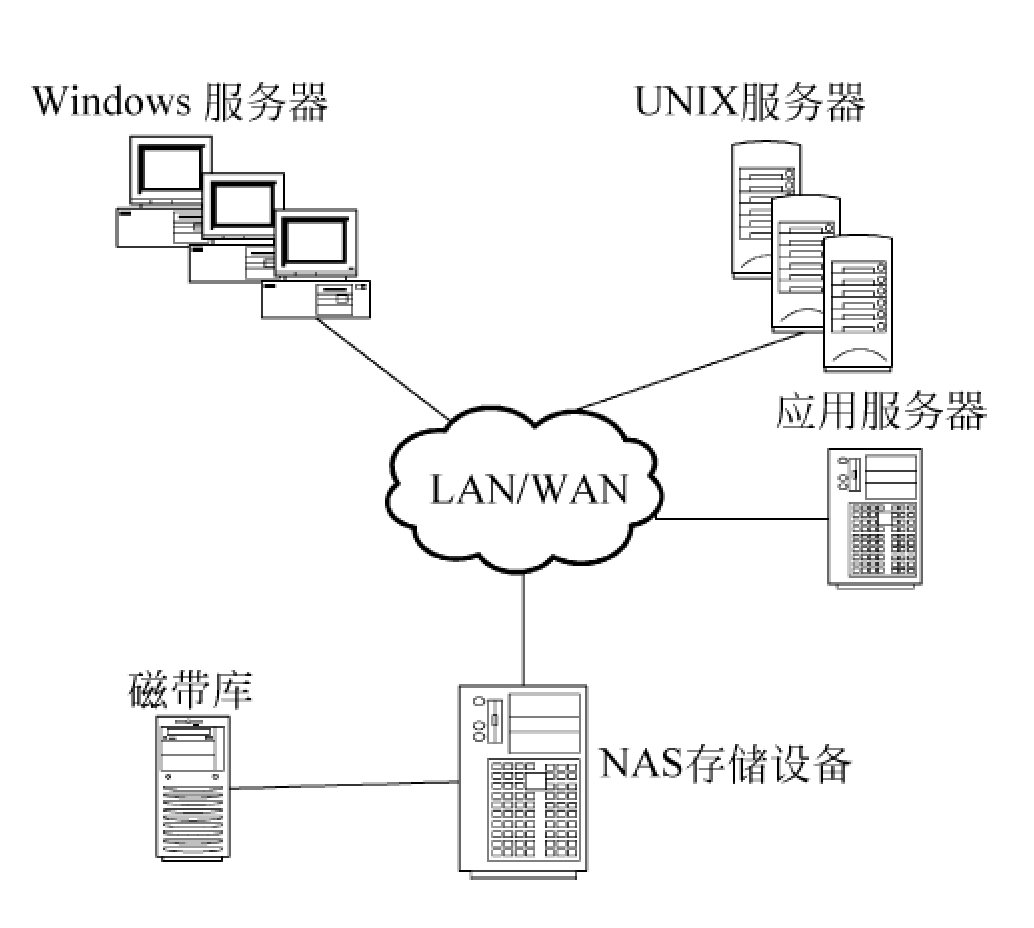


图 6-3　基于 NAS 结构的存储备份

6.2.3　数据备份结构

数据的存储方式直接决定着备份系统的物理构成。目前数据备份结构主要有基于网络附加存储(DAS-Base)备份结构、基于局域网(LAN-Based)备份结构、LAN-Free 备份结构和 SAN-Server-Free 备份结构 4 种。LAN-Based 备份结构针对所有存储类型都可以使用,LAN-Free 备份结构和 SAN-Server-Free 备份结构只能针对 SAN 结构的存储。

1. DAS -Based 备份结构

基于 DAS 的备份系统是最简单的一种数据保护方案,这种备份大多是采用服务器自带的磁带机或备份硬盘,而备份操作也往往通过手工操作的方式进行。因为 DAS 备份系统比较简单,所以比较适用于小型企业用户进行简单的文件备份。DAS-Based 备份结构如图 6-4 所示。

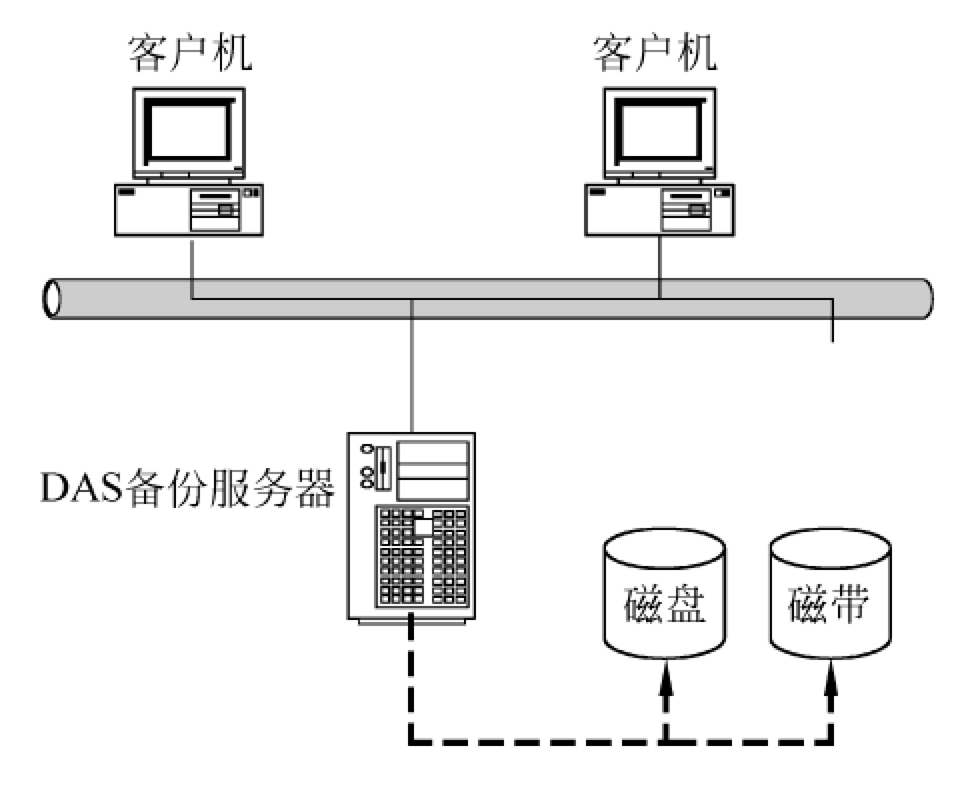


图 6-4　DAS-Based 备份结构

基于 DAS 的备份系统的优点是维护简单,数据传输速度快。它的缺点也很明显,即可管理的存储设备少,不利于备份系统的共享和大型的数据备份要求,而且实时性不高。

2. LAN-Based 备份结构

传统备份需要在每台主机上安装磁带机备

份本机系统，采用LAN-Based备份策略，在数据量不是很大的时候可采用集中备份。一台中央备份服务器将会安装在LAN中，然后将应用服务器和工作站配置为备份服务器的客户端。中央备份服务器接受运行在客户机上的备份代理程序的请求，将数据通过LAN传递到它所管理的、与其连接的本地磁带机资源上。这一方式提供了一种集中的、易于管理的备份方案，并通过在网络中共享磁带机资源提高了效率。

LAN-Based备份结构是小型办公环境最常使用的备份结构，如图6-5所示。在该系统中数据的传输是以局域网络为基础的，首先预先配置一台服务器作为备份管理服务器，它负责整个系统的备份操作。磁带库则接在某台服务器上，当需要备份数据时备份对象把数据通过网络传输到磁带库中以实现备份。

基于LAN-Based备份结构的优点是投资经济、磁带库共享并且能够集中备份管理。它也存在着比较多的缺点，那就是对网络传输压力大，当备份数据量大或备份频率高时，局域网的性能下降很快，所以它不适合大型的网络应用环境。

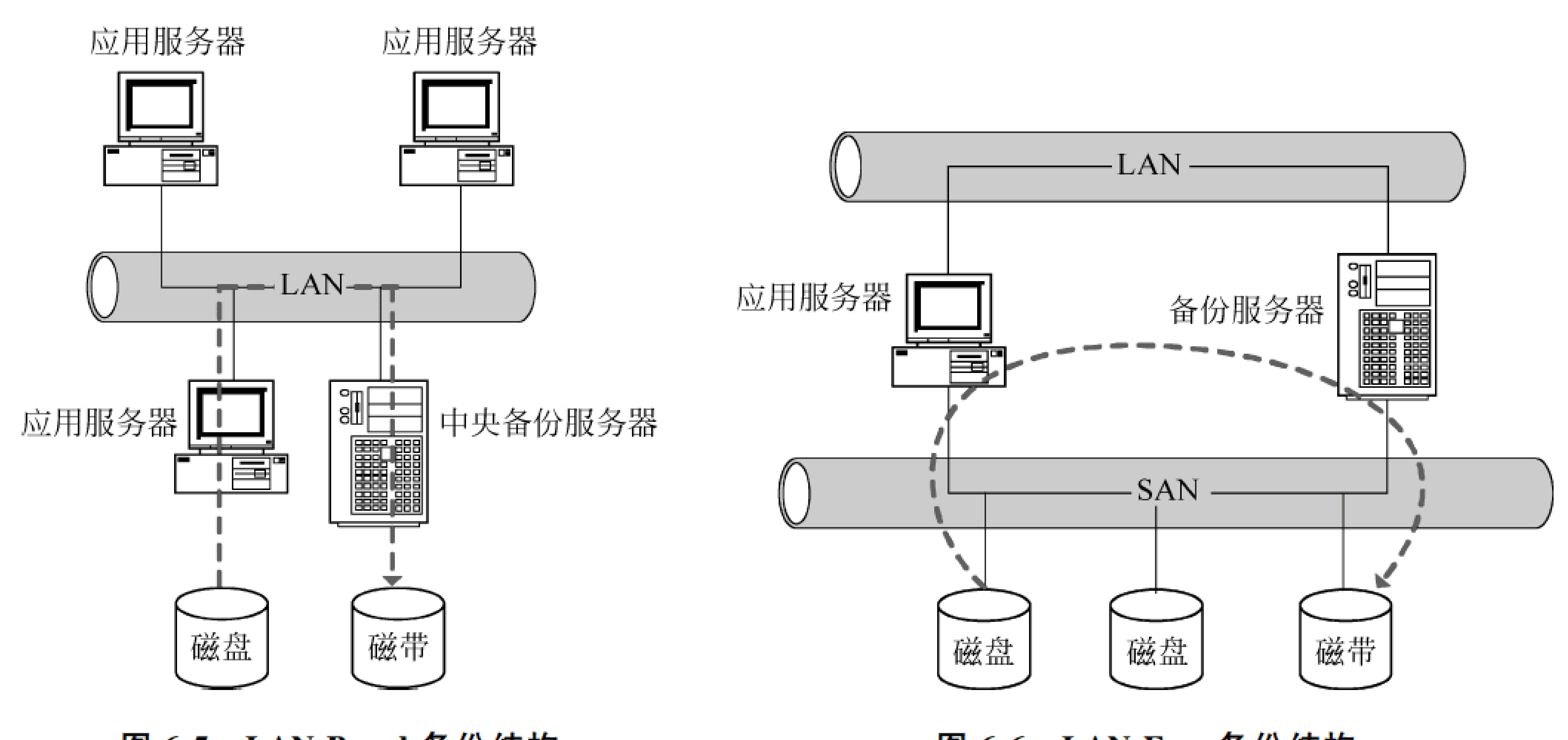


图6-5 LAN-Based备份结构

图6-6 LAN-Free备份结构

3. LAN-Free备份结构

由于数据通过LAN传播，当需要备份的数据量较大、备份时间窗口紧张时，网络容易发生堵塞。在SAN环境下，可采用存储网络的LAN-Free备份结构，需要备份的服务器通过SAN连接到磁带机上，在LAN-Free备份客户端软件的触发下，读取需要备份的数据，通过SAN备份到共享的磁带机，如图6-6所示。这种独立网络不仅可以使LAN流量得以转移，而且它的运转所需的CPU资源低于LAN-Based方式，这是因为光纤通道连接不需要经过服务器的TCP/IP栈，而且某些层的错误检查可以由光纤通道内部的硬件完成。在许多解决方案中需要一台主机来管理共享的存储设备以及用于查找和恢复数据的备份数据库。

尽管LAN-Free备份技术与LAN-Based备份技术相比有很多优点，但LAN-Free备份技术也存在明显不足。首先，服务器仍然参与了将备份数据从一个存储设备转移到另一个存储设备的过程，这在一定程度上占用了服务器的CPU处理时间和服务器内存；另外，LAN-Free备份技术的恢复能力非常依赖用户的应用。

LAN-Free 备份结构的优点是数据备份统一管理、备份速度快、网络传输压力小、磁带库资源共享。缺点是少量文件恢复操作烦琐，并且技术实施复杂，投资较高。

4．SAN-Server-Free 备份结构

另一种减少对系统资源消耗的方法是采用无服务器备份技术。LAN-Free 备份需要占用备份主机的 CPU 资源，如果备份过程能够在 SAN 内部完成，而大量数据流无须流过服务器，则可以极大降低备份操作对系统的影响，SAN-Server-Free 备份就是这样的技术，如图 6-7 所示。

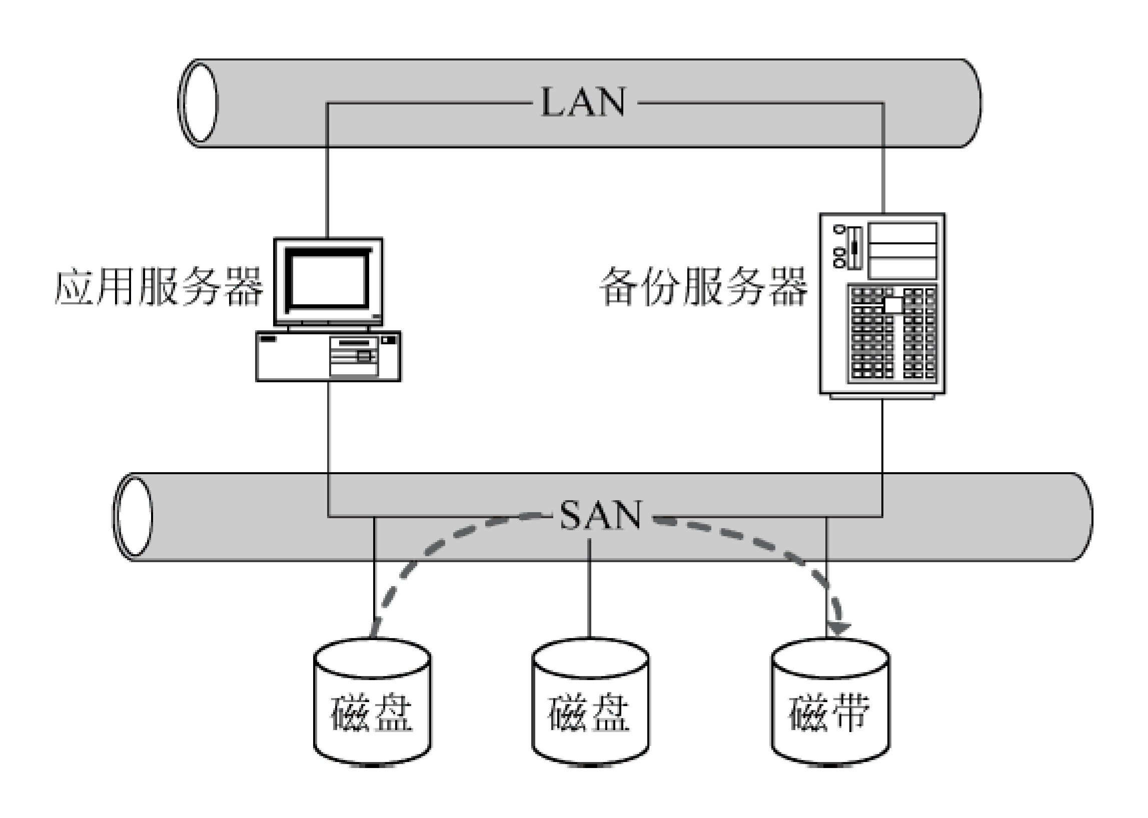


图 6-7　SAN-Server-Free 备份结构

SAN-Server-Free 备份结构如图 6-7 所示，其优点是数据备份和恢复时间短，网络传输压力较小，便于同时管理和备份资源共享。其缺点在于需要特定的备份应用软件进行管理，厂商的类型兼容性问题需要统一，所以实施起来比较复杂，成本也较高，适用于大中型企业进行海量数据备份管理。

目前主流的备份软件(例如 IBM Tivoli、Veritas)，均支持上述备份方案。这 4 种方案均有各自的优点和缺点，在制定备份系统时，用户需要根据数据存储结构及预算等实际情况决定备份系统的物理布局。其中，LAN-Based 备份数据量最小，对服务器资源占用最多，成本最低；LAN-Free 备份数据量居中，对服务器资源占用小一些，但成本相对高一些；SAN-Server-Free 备份方案能够在短时间备份大量数据，对服务器资源占用最少，但成本最高。

6.2.4　数据备份策略

备份策略就是定义备份任务执行的一种方法，即指导备份系统实现备份工作。定义完备份策略后，无须人工进行干涉，备份系统会按备份策略要求备份各类数据。备份策略指确定需备份的内容、备份时间及备份方式。目前被采用最多的备份策略主要有以下 3 种。

1．完全备份(Full Backup)

完全备份是指对某一个时间点上的整个系统或用户指定的所有文件数据进行一次全面的备份。这是最基本也是最简单的备份方式。当发生数据丢失的灾难时，可以只使用一份备份文件快速地恢复所丢失的数据。然而它亦有不足之处，首先它需要备份所有的数据，则工作量大，也需要大量的备份介质；若备份次数较多，则造成大量重复的备份数据；由于需要备份的数据量较大，因此备份所需的时间也就较长。

2. 增量备份(Incremental Backup)

增量备份是指在一次全备份或上一次增量备份后,以后每次的备份只需备份与前一次相比增加或者被修改的文件。在特定的时间段内只有少量的数据发生改变,因此增量备份没有重复的备份数据,既节省了磁盘空间又缩短了备份时间。但在进行恢复操作时,需要查询一系列的备份文件,从最后一次完全备份开始,将记录在一次或多次的增量备份中的改变应用到文件上,需要多份备份文件才可以完成,极大地延长了恢复时间。在这种备份方式下,各盘磁带间的关系就像链子一样,一环套一环,其中任何一盘磁带出了问题都会导致整条链子脱节。

3. 差分备份(Differential Backup)

差分备份是指在一次完全备份后到进行差异备份的这段时间内,对那些增加或者修改的文件进行备份。它的主要目的是将完全恢复时所涉及的备份记录数据限制在两个,以简化恢复的复杂性。差分备份在避免了另外两种策略缺陷的同时又具有了它们的优点:首先,它无须频繁地做完全备份,工作量小于完全备份,因此备份所需的时间短、节省磁盘空间;其次,虽然每次做差分备份工作的任务比增量备份的工作量要大,但是它的灾难恢复相对简单,系统管理员只需要对两份备份文件(完全备份文件和最近一次的差分备份文件)进行恢复就可以将系统恢复。

以上各种备份的数据量不同,按照从多到少的排序为:完全备份>差分备份>增量备份。在恢复数据时需要的备份介质数量也不一样:如果使用完全备份方式,只需上次的完全备份磁带就可以恢复所有数据;如果使用完全备份+增量备份方式,则需要上次的完全备份磁带和上次完全备份后的所有增量备份磁带才能恢复所有数据;如果使用完全备份+差分备份方式,只需上次的完全备份磁带和最近的差分备份磁带就可以恢复所有数据。在备份时要根据它们的特点灵活使用。

在实际应用中,备份策略通常是这 3 种策略的结合。例如,完全备份工作可以在休息日来完成,例如每周日、每月底、每年底等,增量备份或差异备份工作可以夜间进行。还需要注意的是,单盘磁带的容量必须大于或等于所需备份的硬盘容量之和。另外,磁带机本身的性能必须满足备份时间窗口的需求。

6.2.5 数据备份技术

1. 硬件备份技术

硬件备份措施有磁盘镜像、磁盘阵列、双机热备份和双机共享磁盘阵列等。在网络系统中,网络服务器使用率最高且最重要的部分是磁盘系统,磁盘系统的可靠性是服务器至关重要的环节。磁盘系统由磁盘控制卡、SCSI 电缆及硬盘驱动器组成。为了防止磁盘系统出现故障导致系统死机,人们设计了多种方法来保证磁盘系统可靠安全地运行。磁盘双工、镜像及磁盘阵列容错是磁盘系统安全可靠技术的具体实现。下面就对这些技术的应用特点作简单介绍。

1) 镜像技术

镜像技术是在两个或多个磁盘或存储系统上产生同一个数据的镜像视图的一种过程,其中一个存储系统为主镜像系统,另外的存储系统都被认为是从镜像系统。按照镜像存储

系统所处的位置可以分为：本地镜像和远程镜像。本地镜像的主从镜像存储系统是处于同一个 RAID 阵列内，RAID 1 就属于典型的本地镜像技术。远程镜像按请求镜像的主机是否需要远程镜像站点的确认信息，又可分为同步远程镜像和异步远程镜像。

同步远程镜像（同步复制技术）是指通过远程镜像软件，将本地数据以完全同步的方式复制到异地，每一本地的 I/O 操作均需等待远程复制的完成确认信息才能释放。同步镜像使远程复制总能与本地系统要求复制的内容相匹配。当主系统出现故障时，可以很快地切换到从系统，被镜像的远程副本可以保证数目的完整性和可用性。但它存在往返传播造成延时较长的缺点，只限于在相对较近的距离上应用。

异步镜像针对同步镜像的弱点采用了一种完全不同的方式，它允许镜像过程与本地写操作分离，使得本地系统不必由于等待远程存储的写操作而遭受性能上的损耗。镜像完成本地写 I/O 操作后即向应用程序发送完成信号，然后才与异地的镜像代理通信完成数据镜像。异步镜像可以提高本地应用性能，却会产生远程数据的一致性的威胁。远程的复制在任一个指定的时间点上都比源端的数据滞后。这种滞后依赖于网络的带宽和在远程终端上提交写磁盘操作时资源的可用性。当原系统发生错误时，丢失缓冲数据和传输中数据的可能性更大。

2）RAID 技术

RAID 是廉价冗余磁盘阵列。磁盘阵列的提出是保证计算机存储系统可靠性的一个重要发展。RAID 是 UCBerkeley 大学的研究人员 Katz R. H、Gibson G. A 和 Patterson D. A 于 1989 年在《高性能计算的磁盘体系结构》一文中提出来的。RAID 把多块独立的硬盘（物理硬盘）按不同的方式，例如分条（Striping）、分块（Declustering）、交叉存取（Interleaving）等，组合起来形成一个硬盘组（逻辑硬盘），从而提供比单个硬盘更高的存储性能。RAID 技术有多种实现方式，通常采用的有 RAID 0、RAID 1、RAID 3、RAID 5、RAID 10 等。

RAID 0 又称为数据分块，是使用“条”技术来跨越磁盘分配数据的，其目的是将容量和传输率提高到最大，但没有容错功能，一旦硬盘出现故障，阵列中的所有数据将会丢失。

RAID 1 又称镜像法，它使用两个完全相同的盘，即每次将数据同时写入两个盘，一个作为工作盘，另一个作为镜像盘。一旦工作盘发生了致命故障，镜像盘可立即顶上，使系统工作不间断。这种盘阵列可靠性高，但有效容量将减小一半。

RAID 3 即奇偶校验并行交错阵列，每个条带上都有一块空间用来有效存储冗余信息，即奇偶位。奇偶位是数据编码信息，如果某个磁盘发生故障，可以用来恢复数据。即使有多个数据磁盘，也只能使用一个校验磁盘采用奇偶校验的方法检查错误。由于同一个磁盘阵列中，两个或两个以上磁盘同时出现故障的几率很小，所以一般情况下，使用 RAID 3 模式时安全性是可以得到保障的。RAID 3 模式读取数据的速度很快，但写入数据时要计算校验位来获知写入的校验磁盘，因此写入速度相对较慢。

RAID 5 是一种旋转奇偶校验独立存取阵列，它按一定规则把奇偶校验信息均匀分布在阵列中所有的盘上，是一种容错能力分布合理的阵列。为了提供冗余，它最少需要 3 个磁盘（不包括热备份盘）。RAID 5 是通常使用最多的数据保护方案。

RAID 10 实际上是 RAID 0＋RAD 1。它采用分块和镜像技术，通过分块镜像集实现。采用分块技术，多个磁盘可并行读写，磁盘 I/O 性能很高；采用镜像存储使得可靠性是所有磁盘阵列中最高的。由于集中了 RAID 0 和 RAID 1 的优点，RAID 10 的性能是所有 RAID

类型中最好的，但代价较高。

这几种常用的RAID技术的特征比较如表6-1所示。

表6-1 常用的RAID模式的特征比较

RAID模式	容错性	冗余类型	热备份选择	磁盘数据	有效磁盘容量
RAID 0	没有	没有	没有	一个或多个	磁盘的总容量
RAID 1	有	复制	有	偶数个	磁盘的总容量的50%
RAID 3	有	奇偶校验	有	3个或3个以上	磁盘的容量的$(n-1)/n$
RAID 5	有	奇偶校验	有	3个或3个以上	磁盘的容量的$(n-1)/n$
RAID 10	有	复制	有	只需4个	磁盘总容量的50%

3）磁盘双工

磁盘双工就是在一台服务器内采用两个磁盘控制器，各自接一个性能相同的硬盘。在系统工作时，将数据同时存入两个硬盘，当一个硬盘或一块控制器出现故障时，可以继续使用另一个磁盘系统，这样就能实现确保网络正常运行。双工的传输速度比较快，但成本较高。

4）双机热备份

双机热备份又叫双机容错，就是配置两台完全一致（也可不一致）的服务器系统。一台作为主服务器，另一台作为备份服务器。两台服务器上安装高速镜像卡或普通的100Mbps网卡，通过高速链路（例如光纤或专用电缆）连接起来，系统运行时，数据在存入主服务器的同时也存入备份服务器。也就是说备份服务器完成与主服务器同样的操作，当主服务器运行出现故障时，系统控制权切换到备份服务器，即备份服务器立即代替主服务器运行，实时保证网络系统不中断。当主服务器系统修复后，控制权需再切换回到主业务系统，使双机系统恢复正常冗余工作模式。双机热备份可以防止单台计算机的物理损坏，但无法防止逻辑损坏。

磁盘镜像与磁盘阵列不同的地方在于磁盘阵列可以防止多个硬盘出现故障，而磁盘镜像只能防止单个硬盘的物理损坏。双机热备份和磁盘阵列系统是完备的硬件容错系统，可防止整机出现故障。这4种措施都属于在线的硬件级备份，对火灾、水淹、线路故障造成的系统损坏和逻辑损坏都无能为力。实际上只有离线的、远离运行中心并妥善保管的备份才会比较可靠，这样的备份才可用作灾难性恢复。

2. 软件备份技术

软件备份技术是通过操作系统提供的备份软件或专业备份软件将系统数据复制到可以异地存放的存储介质上。软件备份需从3个方面考虑：选择合适的备份存储介质；备份软件的选择；制定合适的备份策略。其中处于核心地位的是备份软件的选择。

通常备份软件分为静态备份软件和动态备份软件两类。静态备份软件：能够方便地选择备份内容，但不能定时自动备份，如果实现自动备份，还要自己编写脚本文件或使用操作系统的计划任务之类的功能。动态备份软件：能够实现选择备份时间、自动后台作业、定时完成操作等功能。一个好的备份软件应该具有如下优点：

（1）安装方便、界面友好、使用方便灵活。

（2）支持多系统、多平台、多文件格式的备份。

(3) 支持文件打开状态备份。

(4) 支持在网络中的远程集中备份。

(5) 支持备份介质自动加载的自动备份。

3. 冷备份技术与热备份技术

按照备份数据的在线状态和备份的实时性,还可分为冷备份技术和热备份技术。

1) 冷备份技术

冷备份又叫离线备份,它是指当执行备份操作时,服务器将不接受来自用户和应用程序对数据的更新。离线备份很好地解决了在备份操作进行时更新数据带来的数据不一致性问题,且备份速度快。但在实施备份的全过程中,服务器只能做备份而不能及时响应用户的需求,用户需要等待很长时间。

2) 热备份技术

热备份也称在线备份,即同步数据备份,就是用户和应用程序正在更新数据时,系统也可以进行备份。由于备份与系统同步,则资源占用比较多,但恢复时间非常短。在热备份中,数据有效性和完整性是一个很大的问题,如果备份过程中产生了数据不一致性,会导致数据的不可用。解决此问题的方法是对于一些总是处于打开状态的重要数据文件,备份系统可以采取文件的单独写/修改权,保证在该文件备份期间其他应用不能对它进行更新。热备份的技术主要有两个:写前复制和软件快照技术。

写前复制大多数在数据库备份环境下实现,是指当正在备份的数据库对象发生改变时,将磁盘上的原有数据块复制到一个临时磁盘位置,并使用一个特殊的位图索引标明原有块的位置以及临时存储的相对位置。当数据库对象结束备份时,就清除位图索引,释放临时存储的数据块,提供给下一次使用。

软件快照技术与写前复制技术相类似,是在镜像磁盘上建立第三次复制的一种方法。快照可以在软件中建立,提供文件系统和数据库的即时映像,这样当备份的时候就可以获得完整的数据复制。软件快照将每一个文件系统或数据库的存储块都保存其存储分配的一份复制。文件系统和数据库的视图就由这些即时的块分配所决定,所以在任何一个时刻,假如希望能取得文件系统映像,那么就需要保证对这些块的可访问性,这就是软件快照。软件快照能冻结文件系统的块分配视图,当然也可以冻结系统的子集,如目录或数据库的表。

4. 其他备份技术

根据备份设备与系统的相对位置还可分为本地备份技术和远程(异地)备份技术;按照数据备份的网络实现方式,可分为单机备份技术和网络备份技术;根据备份的自动化程度,还可分为手工备份和自动备份等。

6.2.6　数据备份软件

目前国内外对数据备份研究越来越多,备份软件产品也丰富多样,用户需要根据备份技术的需求及成本预算等选定备份产品。下面将对国内外几种常用的数据备份产品进行介绍。

1. 国外产品

国外对数据备份系统的研究和开发,开始于 20 世纪 80 年代中期。到目前为止,成熟的

产品主要集中在少数知名 IT 公司手中,例如 IBM、VERITAS、EMC 等,这些在存储和数据备份恢复领域处于领导地位的公司都有各自的存储备份技术和优秀的数据备份恢复系统。

1) IBM 公司的异地备份系统

IBM 的 PPRC(Peer to Peer Remote Copy,点对点远程复制)技术是基于 ESS 企业级数据存储服务器,通过 ESCON(Enterprise Systems Connection,企业管理系统连接,一种光纤通道)通道建立配对的逻辑卷备份技术。这是 IBM 的最高级别备份系统,主要适用于大、中型和电信企业。PPRC 备份系统能够自动将源卷上的数据同步到目标卷,实现以存储为基础的、实时的、与应用无关的数据远程镜像功能,可根据需要选择同步或异步方式。PPRC 实现较为简单,纯粹基于硬件,是无数据丢失且具有完全恢复功能的灾难恢复系统。

2) VERITAS 公司多层次灾难恢复系统

VERITAS 提供的是基于软件的备份复制的多层次灾难恢复系统。这个系统集成了 Volume Replicator、Cluster Server 和 Global Cluster Manager,其代表是 VERITAS Volume Replicator。VERITAS Volume Manager 可以在不同的地理位置建立数据集的镜像,以可靠、高效、一致的方式通过 IP 网络将数据复制到异地站点,无论企业的存储是分布于 LAN、MAN 还是 WAN 层次上。在具体的功能配合上,VERITAS 的产品各司其职:VERITAS Cluster Server 处理本地可用性问题;VERITAS Volume Replicator 将重要数据复制到远端站点;Global Cluster Manager 监控并管理每个站点的复制工作和集群。如果出现站点故障或主站点的应用完全失效,Global Cluster Manager 将控制并转移生产任务到备用站点,重新引导客户机的流量。

3) EMC 公司的 SRDF 同步复制系统

EMC 的 SRDF(Symmetrix Remote Data Facility,Symmetrix 远程数据镜像技术)实现了数据在不同环境间的实时有效复制。SRDF 是在 Symmetrix 成功经验的基础上,通过对磁盘子系统的性能不断改进而产生的。SRDF 拥有两套磁盘子系统,分别称为 R1 和 R2,存放实时数据复制的 R2 子系统被安置在与存放原始数据复制的 R1 子系统不同的地点。这样就确保了在数据中心发生故障时,R2 系统仍然是可用的,而且与 R1 是同步的。由于使用了不同的子系统,所以可对 R1 和 R2 分别进行寻址。第二个数据复制 R2 可以按照只读模式供附加在第二个子系统上的第二个处理器使用。总的来说,SRDF 磁盘子系统被分为本地 R1 和远程 R2 两部分,任何写入 R1 的数据都会同时被自动复制到 R2 之上。SRDF 的工作独立于操作系统,为用户提供了一种稳定良好的远程镜像备份系统,并且具有很大的灵活性和可控性。

2. 国内产品

国内的数据备份产品仍处于初期发展阶段,我国的网络数据安全备份技术起步于 20 世纪 80 年代后期,取得了一系列研究成果,也形成了一些产品,例如优备 EUBASE、全球盾、爱数等。

1) 联想一键恢复 LEOS 系统

联想集团的一键恢复 LEOS 系统主要用来保护用户操作系统,具有数据隐藏分区模块、硬盘保护模块和一键恢复模块。用户可以轻松地将系统数据恢复到最近一个时间点备份。

2）爱数数据备份方案

作为国内首个具有完全知识产权的系统备份与灾难恢复产品提供商，爱数率先提供当前最先进的基于磁盘的完整存储备份方案及企业在线备份服务。主要产品有爱数备份存储柜、爱数备份软件企业版和爱数备份卫士。爱数的产品已应用于上海海鼎信息工程股份有限公司、北京视翰科技等多家企业。

国内的数据备份系统产品与国外公司的产品相比，市场占有率低，知名度低。同时，在产业规模、技术水平、开发能力和国际竞争能力等方面与国际先进水平相比有着很大的差距。为应对国际品牌的挤压和发展民族品牌，研究和开发数据备份系统变得更加需要和迫切。

6.3　数据恢复的基础知识

数据最根本是存储在介质中的比特符号，因此要对数据进行恢复，就必须从介质中入手。介质有硬盘、软盘、光盘、U盘、数码卡和RAID等，而最常用的介质就是硬盘，所以数据恢复一般指的是硬盘数据恢复。要做到恢复数据，就不得不了解硬盘的数据结构、文件的存储原理和操作系统的启动流程等基础知识。

6.3.1　硬盘的基础知识

1. 硬盘的逻辑结构

1）盘片

硬盘的盘片是硬质磁性合金盘片，一些高速硬盘也可能采用玻璃做基片，其直径主要有1.8英寸、2.5英寸和3.5英寸等。每个盘片都有两个盘面，每个盘面都可以存储数据，按顺序从上而下自0开始依次编号。

2）磁道

磁盘在格式化时被划分为许多同心圆，这些同心圆轨迹叫做磁道(Track)，从外向内自0开始编号。信息以脉冲串的形式记录在这些轨迹中，这些同心圆不是连续记录数据，而是被划分成一段段的有着相同角速度的圆弧。

3）扇区

每个磁道按512B为单位被划分成一段段的圆弧，每段圆弧叫做一个扇区，扇区从1开始编号，每个扇区中的数据作为一个单位同时读出或写入。一个扇区有两个主要部分：存储数据地点的标识符和存储数据的数据段。

4）柱面

所有盘面上的同一磁道构成一个圆柱，通常称为柱面，每个圆柱上的磁头，由上而下从0开始编号。数据的读写是按柱面进行的，一个磁道写满数据，不是在同一盘面的下一磁道来写，而是在同一柱面的下一个盘面来写。一个柱面写满后，才移到下一个柱面，从下一个柱面的1扇区开始写数据。

硬盘的容量由盘片数(磁头数)、柱面数和扇区数决定，其计算公式为：

$$硬盘容量 = 盘片数 \times 柱面数 \times 扇区数 \times 512B$$

2. 硬盘总体存储结构

整个硬盘的总体结构由主引导扇区、各个分区和未划分空间(如果存在)组成。

主引导扇区位于整个硬盘的最前端,0柱面0磁道1扇区,也就是LBA 0扇区。它占用一个扇区,里面包含MBR(主引导记录)和磁盘分区表(DPT)两大部分。在主引导扇区的后面是62个保留扇区,是未用的空白区域。从第63个扇区开始(0柱1道1扇)后面的所有空间开始用于分区使用。

硬盘的总体存储结构如图6-8所示。

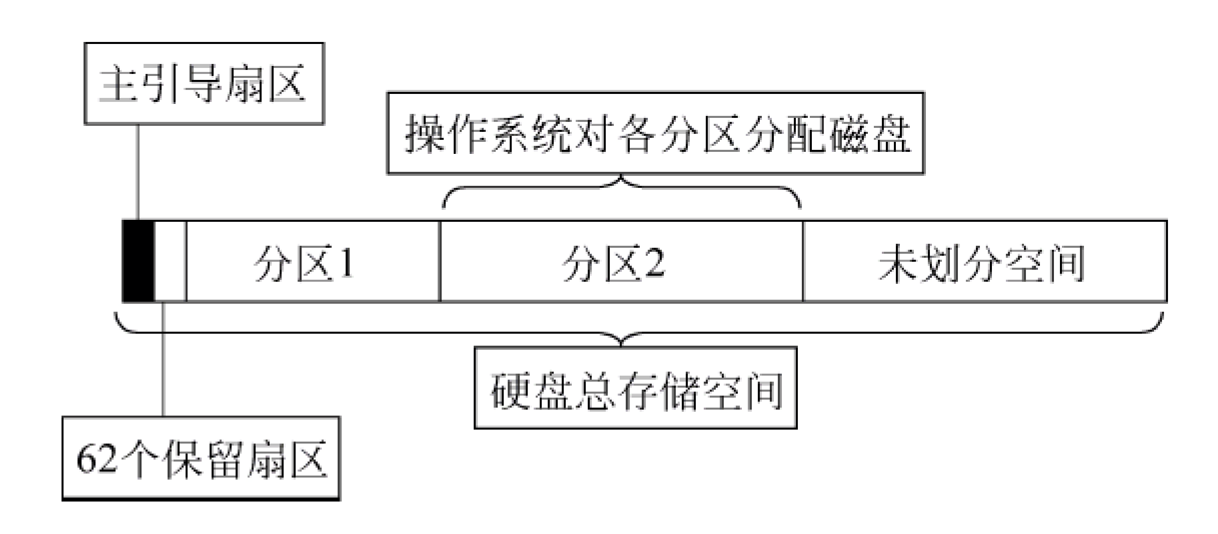


图6-8 硬盘的总体存储结构

3. 硬盘的数据结构

刚生产出来的硬盘是无法使用的,若要使用就先将它分区、格式化,然后再安装上操作系统才可以使用。而在这一过程中,要将硬盘分成主引导区(MBR)、操作系统引导记录区(DBR)、FAT表、DIR目录区和DATA数据区5个部分。

1) MBR

MBR(Main Boot Record,主引导区)位于整个硬盘的0磁道0柱面1扇区,包括硬盘主引导记录(MBR)和分区表(DPT)。主引导记录的作用就是检查分区表是否正确以及确定哪个分区为引导分区,并在程序结束时把该分区的启动程序(也就是操作系统引导扇区)调入内存加以执行。分区表以80H或00H为开始标志,以55AAH为结束标志,共64B,位于本扇区的最末端。主引导记录由fdisk命令创建分区时建立的,它不依赖于任何操作系统。

2) DBR

DBR(DOS Boot Record,操作系统引导区)通常位于硬盘的0磁道1柱面1扇区,是操作系统可以直接访问的第一个扇区,它包括一个引导程序和一个被称为BPB(BIOS Parameter Block)的分区参数记录表,最后的结束标志为55AA。BPB参数块记录着本分区的起始扇区、结束扇区、文件存储格式、硬盘介质描述符、根目录大小、FAT个数,分配单元的大小等重要参数。DBR由format程序建立,不同的操作系统,DBR的结构也会不同。

3) FAT

FAT(File Allocation Table,文件分配表)是操作系统的文件寻址系统。为了防止意外损坏,FAT一般做两个,第二个FAT为第一个FAT的备份。同一个文件的数据并不一定完整地存放在磁盘的一个连续的区域内,而往往会分成很多段,像一条链子似的存放,这样存放更多是为了读写速度的考虑。在FAT表中,00表示未分配的簇;FFF7是坏扇区

标记 Bad；FFFF 是文件的最后一个簇，即文件结束标记 Eof；其他数字表示文件已分配的簇号。

4）DIR

DIR(Directory，根目录区)紧接在第二个 FAT 表之后，DIR 记录着每个文件(目录)的起始单元、文件的属性等。定位文件位置时，操作系统根据 DIR 中的起始单元，结合 FAT 表就可以知道文件在磁盘的具体位置及大小了。

5）DATA

DATA(数据区)在 DIR 区之后，是数据存储区，虽然占据了硬盘的绝大部分空间，但若没有前面各区，它是没有任何意义的。通常所说的格式化程序(指高级格式化，例如 DOS 下的 format 程序)并没有把 DATA 区的数据清除，只是重写了 FAT 表而已，至于分区硬盘，也只是修改了 MBR 和 OBR，绝大部分的 DATA 区的数据并没有被改变，这也是许多硬盘数据能够得以修复的原因。

6.3.2　文件的存储原理

数据以文件形式存储，要对数据进行恢复，就必须了解文件的存储原理。

1. 文件的读取

操作系统从目录区中读取文件信息(包括文件名、后缀名、文件大小、修改日期和文件在数据区保存的第一个簇的簇号)。这里假设第一个簇号是 0023。操作系统从 0023 簇读取相应的数据，然后再找到 FAT 的 0023 单元，如果内容是文件结束标志 FFFF，则表示文件结束，否则内容保存数据的下一个簇的簇号，这样重复下去直到遇到文件结束标志。

2. 文件的写入

当要保存文件时，操作系统首先在 DIR 区中找到空区写入文件名、大小和创建时间等相应信息，然后在 DATA 区找到闲置空间将文件保存，并将 DATA 区的第一个簇写入 DIR 区，其余的动作和上边的读取动作差不多。

3. 文件的删除

Windows 9x 的文件删除工作却是很简单的，仅是将目录区的文件的第一个字符改成了 E5 就表示将该文件删除了。

6.3.3　操作系统的启动流程

各种不同的操作系统启动流程不尽相同，这里以 Windows 9x/DOS 的启动流程为例。

第 1 阶段：系统加电自检 POST 过程。POST 是 Power On Self Test 的缩写，也就是加电自检的意思，计算机执行内存 FFFF0H 处的程序(这里是一段固化的 ROM 程序)，对系统的硬件(包括内存)进行检查。

第 2 阶段：读取分区记录和引导记录。当计算机检查到硬件正常并与 CMOS 设置相符后，按照 CMOS 设置从相应设备启动(这里假设从硬盘启动)，读取硬盘的分区记录(DPT)和主引导记录(MBR)。

第 3 阶段：读取 DOS 引导记录。计算机正确读取分区记录和主引导记录后，如果主引

导记录和分区表校验正确，则执行主引导记录并进一步读取 DOS 引导记录（位于每一个主分区的第一个扇区），然后执行该 DOS 引导记录。

第 4 阶段：装载系统隐含文件。将 DOS 系统的隐含文件 Io. sys 载入内存，加载基本的文件系统 FAT，这时候一般会出现 Starting Windows 9x 的标志，Io. sys 将 Ms. sys 读入内存，并处理 System. dat 和 User. dat 文件，加载磁盘压缩程序。

第 5 阶段：实 DOS 模式配置。系统隐含文件装载完成，计算机将执行系统隐含文件，并执行系统配置文件（Config. sys），加载 Config. sys 中定义的各种驱动程序。

第 6 阶段：调入命令解释程序（Command. com）。系统装载命令管理程序，以便对系统的各种操作命令进行协调管理（用户所使用的 dir、copy 等内部命令就是由 Command. com 提供的）。

第 7 阶段：执行批处理文件。计算机将一步一步地执行批处理文件中的各条命令。

第 8 阶段：加载 Win. com。Win. com 负责将 Windows 下的各种驱动程序和启动执行文件加以执行。至此启动完毕。

6.4 硬盘数据恢复技术

造成数据丢失的原因非常多，每种情况都有特定的症状出现，或者多种症状同时出现，一般情况下，只要数据区没有被覆盖，都是可以恢复的。下面列出一些能够进行数据恢复的前提条件：

(1) 数据有进行备份，例如有两份 FAT 表。

(2) 数据的实际有效性：数据不一定丢失了，可能只是操作系统找不到，而从物理扇区上来看它仍然是存在的。例如文件删除，事实上，只是将 FAT 表中文件名的首字节改为 0E 而已，此时文件依然存在。

(3) 数据本身为标准数据：例如引导扇区、DLL 文件等。

(4) 数据本身可以由其他信息统计再生。

(5) 破坏的程度较小：误删除文件、误分区以及误格式化等都不会彻底破坏数据，只有低格和扇区覆盖操作才会彻底破坏数据。

6.4.1 主引导区的恢复

对于开机自检后提示 Miss operation system，而且 DOS 下查看 C 盘内容完整，这是属于主引导区故障。另外在计算机启动中，系统能够通过自检并检测到硬盘，但在即将进入操作系统之前提示 Disk boot failure insert system disk and press enter，这也是主引导区错误。

对于这一类故障，可以使用 fdisk/MBR 或 fixmbr/MBR 命令来解决。fdisk/MBR 可以进行无条件重写主引导区，而且可以保留原有的数据。fdisk/MBR 堪称是对付硬盘在 BIOS 下可以识别而 DOS 下无法操作的第一工具。fixmbr/MBR 是微软提供的一个在使用 Windows 2000 恢复控制台时，专门修复 MBR 的修复程序，该工具通过全盘搜索来决定硬盘分区，并重新构造主引导扇区。

6.4.2　分区表的恢复

如果使用 PartitionMagic 时操作失误，导致无法进入系统或者是进入系统后文件打不开等情况，即是典型的分区表故障。若分区表受损，就需要重建分区表，可通过相应的工具软件进行修复。

DiskGenius 是一款国产的硬盘分区软件，采用全中文图像界面，以图表的方式将硬盘分区表的详细结构表示出来，操作简单。如果硬盘分区表被分区调整软件或病毒破坏，引起硬盘和系统瘫痪，DiskGenius 可通过未被破坏的分区引导记录信息重建分区表。

Partition Doctor 是一款运行于系统下的软件，所以只能在系统正常的情况下才能起作用。一般是在一个运行正常的系统之下挂上待修理的硬盘，运行 Partition Doctor 进行手动修理。

DiskEdit 是一款只能运行 DOS 下的工具软件，可以直接编辑磁盘的任意扇区。

WinHex 是目前使用最多的一款工具软件，有着完善的分区管理能力和文件管理功能，能自动分析分区链和文件簇链，并能以不同的方式进行不同程度的备份，直至克隆整个硬盘。它能够完整地显示和编辑任何一种文件类型的二进制内容，其磁盘编辑器可以编辑物理磁盘和逻辑磁盘的任一扇区，内存编辑器可以直接编辑内存。可以说，它是目前功能最强大的软件之一，是系统维护的最好工具。

6.4.3　DBR 的恢复

对于一台计算机，在 Windows 系统下打开一个分区时提示未被格式化，并且在 DOS 下进入此分区时提示 General Fail Reading Drive，在这种情况下，如果使用格式化工具对所要访问的分区进行格式化，就能很轻松地进入此分区，但代价就是此分区下所有的数据都将不复存在。这就是 DBR 受损，若要解决此问题，就需要恢复 DBR。

高版本的格式化工具在格式化分区时，一般都会在第六扇区对 DBR 做个备份，也可以自己做备份，若这个备份完好无损，则可以直接使用这个备份修复损坏的 DBR。如果备份的 DBR 也损坏了，就得从相同文件系统的分区里复制一个 DBR，复制完 DBR 后需要进行相应的参数修改，因为不同的分区其 FAT 表长度、簇大小、分区长度等参数都不相同。如果不对这些参数进行修改，虽然能访问该分区，但其文件系统的参数不正确，不能正确访问文件，所以必须把这些参数修改为符合实际情况的参数。

现在有很多恢复软件，例如 NDD、DiskMan 等，它们对 DBR 的恢复是建立在备份 DBR 完好的基础上的。若 DBR 损坏了，就只能通过手工进行恢复。

6.4.4　FAT 表的恢复

FAT 表记录着硬盘数据的存储地址，每一个文件都有一组 FAT 链指定其存放的簇地址。FAT 表的损坏意味着文件内容的丢失。但好在 DOS 系统本身提供了两个 FAT 表，如果目前使用的 FAT 表损坏，可用第二个进行覆盖修复。但由于不同规格的硬盘其 FAT 表的长度及第二个 FAT 表的地址也是不固定的，所以修复时必须正确查找其正确位置。可用 Debug 的 m 命令来将第二个 FAT 表移到第一个 FAT 表处，也可以通过 Norton 8.0 中的 ndd.exe 来很方便地恢复损坏的 FAT 表。

若第二个 FAT 表也损坏了，则无法通过备份 FAT 表来恢复，但若文件数据仍然存放在硬盘数据区中，还是可以采用相关的修复法来挽救数据的。通过 Scandisk.exe 程序可以找回丢失 FAT 链的扇区数据，如果是文本文件则可从中提取并组合成完整的文件，只要将文件名改过来就行了；如果是二进制的数据文件，则很难恢复出完整的文件。

6.4.5 文件误删除的恢复

一般在删除文件的时候，先是把文件转移到回收站，这只是对文件进行了逻辑删除，并没有将文件彻底删除。如果想把删除的数据找回来，通过回收站还原就可以了。但如果在删除的时候按住了 Shift 键或在回收站中选择彻底删除，那么就相当于把数据进行了物理删除，这时候就无法通过一般的方法来找回删除的数据了，但通过数据恢复软件可以达到修复误删除文件的目的。

EasyRecovery：由著名的数据恢复公司 Ontrack 推出，使用 Ontrack 公司复杂的模式识别技术找回分布在硬盘上不同文件的碎块，并根据统计信息对这些文件碎块进行重整。接着 EasyRecovery 在硬盘中建立一个虚拟的文件系统并列出所有的文件和目录。哪怕整个分区都不可见或者硬盘上只有非常少的分区维护信息，EasyRecovery 仍然可以高质量地找回文件。

FinalData：最大优势是恢复速度快，可以减少搜索丢失数据所需要的漫长的等待时间。不仅恢复速度快，而且在数据恢复方面功能十分强大，不仅可以按照物理硬盘或者逻辑分区来进行扫描，还可以通过对硬盘的绝对扇区来扫描分区表，找到丢失的数据。在 Windows 环境下删除一个文件，只要目录信息没从 FAT 或者 MFT(NTFS)删除，就意味着文件数据仍然留在磁盘上，所以从技术角度来讲恢复是可行的。FinalData 就是通过这个机制来恢复丢失的数据的。另外，FinalData 可以很容易地对格式化后的文件或被病毒破坏的文件进行恢复。甚至在极端的情况下，例如目录结构被部分破坏也可以恢复，只要数据仍然保存在硬盘上，FinalData 都可以将数据恢复过来。

但如果文件在删除之后，其存储的磁盘空间进行过写操作，那在通常情况下恢复的几率为零。因此，误删除文件可以恢复的重要前提就是不要在删除文件所在的分区进行写操作。

6.4.6 磁盘坏道的处理

由于硬盘使用时间过长或操作人员的失误，硬盘可能会出现各种各样的问题，而硬盘坏道便是这其中最常见的问题。硬盘坏道分为逻辑坏道和物理坏道两种，前者是由于操作人员的软件操作失误造成的，可用软件修复；后者是硬盘磁道上的物理损伤，只能通过更改硬盘分区或扇区的使用情况来解决。在这里硬盘和磁盘是同一个概念。

1. 磁盘坏道的现象

如果硬盘一旦出现下列这些现象，就说明硬盘可能出现了坏道：

(1) 当打开、运行或复制某个文件时，硬盘出现操作速度变慢，且有可能长时间操作不成功或反复读某一区域，或出现异响，或 Windows 系统提示“无法读取或写入该文件”，这些都可表明硬盘可能出现了坏道。

(2) 若每次开机时，Scandisk 磁盘扫描程序都会自动运行，这表明硬盘上有需要修复的重要错误，例如坏道。在运行该程序时若不能顺利通过，表明硬盘肯定有坏道。若扫描虽然

可通过，但会出现红色的"B"标记，表明硬盘也有坏道。

（3）计算机启动时硬盘无法引导，用软盘或光盘启动后可看见硬盘盘符但无法对该区进行操作或操作有误或看不见盘符，都表明硬盘上可能出现了坏道。具体表现如开机自检过程中屏幕提示 Hard disk drive failure、Hard drive controller failure 或类似信息，则可以判断硬盘驱动器或硬盘控制器有硬件故障；读写硬盘时提示 Sector not found 或 General error in reading drive C 等类似错误信息，则表明硬盘磁道出现了物理损伤。

（4）计算机在正常运行中出现死机或"该文件损坏"等问题，也可能和硬盘坏道有关。

2. 逻辑坏道的修复

对于非系统区的逻辑坏道，可以在系统启动后，借助 Windows 下的磁盘扫描工具，对需要修复的盘符进行完全扫描，并选择自动修复错误，则可修复逻辑坏道。如果逻辑坏道出现在系统区而导致无法正常启动，可以使用 Windows 98/Me 的启动盘，在 DOS 提示符下输入"Scandisk C"，一旦发现坏道，程序会提示修复，选择修复即可。

3. 物理坏道的修复

硬盘的物理坏道具有传染性，因为对有坏道的硬盘进行扫描，就会对硬盘的物理坏区强制进行多次读写，会使坏道向周边扩散，使硬盘上更多的数据处于危险的境地。因此，当发现硬盘有物理坏道后，不要对其进行扫描，而是要把已有的坏道标记隐藏，以后不再对这块区域进行读写操作，以免坏道扩散。

对于如何处理硬盘中的坏道，可以用 Scandisk 磁盘检测、修复命令和相关的磁盘工具软件进行，例如 PartitionMagic 或 DiskMan 等。处理时，首先要估计坏道所处的位置，可用 Scandisk 计算得到。当 Scandisk 在查到坏道停止时，注意观察 Scandisk 停止时的数值，例如 22%。假设硬盘总容量为 2GB，2GB×22%＝0.44GB，这说明硬盘坏道出现的起始位置大致为 440MB 处。当在标记硬盘坏道时，须考虑到硬盘坏道易向周边扩散，应该将坏道区域的范围增大，例如留下 40MB 作为坏道缓冲区。那么，已知的 400MB 的正常硬盘可作为第一个分区容量，其余 1.6GB 按 200MB 为单位分为 8 个区，再使用 Scandisk 检查所有分区，将无法通过 Scandisk 检测的分区利用磁盘工具删除或隐藏，以确保系统不再读写这些区域，其余相邻的分区可合并后使用。硬盘的删除、隐藏、分区或合并都可用磁盘工具进行。

还可联合使用 fdisk 命令和 format 命令来实现。跟上述方法一样，首先要已知硬盘容量，例如 2GB，对有问题的硬盘先用 fdisk 命令分成一个盘，再用 format 命令对其进行格式化。当格式化碰到无法修复的坏道时，记录下进行的百分比，例如 22%，然后按 Ctrl＋Break 组合键强行终止格式化。这说明坏道出现在 440MB 之处，用 fdisk 命令将 400MB 划为一个逻辑盘，40MB 作为缓冲区。之后，再将后面的硬盘估计出坏道的大概范围，例如 10%，将这部分坏道也划为一个小逻辑盘，这个小盘不用格式化。在总格式化完成后，将这个小盘删除，就能将坏道隐藏了。

还有一类特殊的硬盘坏道——0 磁道损坏。当 0 磁道出现损坏时，系统会显示"Track 0 bad，diskunusable"，即"零磁道损坏，硬盘无法使用"或用磁盘扫描程序扫描时其 0 扇区出现红色"B"。当出现这个损坏时，一般人往往将出现这样故障的硬盘作报废处理，其实合理运用一些相关的磁盘软件，把损坏的 0 扇区隐藏，而用 1 扇区代替，就可解决此问题。相关的工具软件有 PartitionMagic、DiskEdit、SmartEdisk、DiskMan 等。

思 考 题

(1) 什么是数据恢复和数据备份？简述数据恢复与数据备份的关系。
(2) 造成数据危险的原因有哪些？
(3) 为什么要数据备份？数据备份的原则和要求是什么？
(4) 数据备份有几种模式？各有何特点？
(5) 数据备份的结构有几种？
(6) 数据备份的策略有几种？各有何特点？
(7) 有几种数据备份技术？
(8) 如何制定数据备份策略？
(9) 硬盘的物理和逻辑结构分别是什么？
(10) 系统中文件是如何存储的？
(11) 当备份 FAT 表损坏时，怎样进行 FAT 表的恢复？
(12) 磁盘坏道了会出现怎样的现象？如何进行磁盘坏道情况下的数据恢复？

参 考 文 献

[1] 张敏情，杨晓元. 数据安全基础. 北京：人民邮电出版社，2008.
[2] 王文珍，唐红文，张成利. 基于磁盘的数据备份与恢复管理系统的研究. 物理装备，2007，17(4)：276～279.
[3] 黄晶. 数据备份系统的研究与实现. 华中科技大学硕士学位论文，2008.
[4] 伍江江. 支持服务恢复的数据备份恢复技术研究与实现. 国防科技大学硕士学位论文，2007.
[5] 丁晓红. 网络环境下信息数据安全备份的实施方案. 电子机械工程，2004，(1)：40～41.
[6] 颜洪梅，战守义，杨方廷. 网格系统中数据备份技术的研究. 计算机应用研究，2006：(12)154～156.
[7] 石青，陈锋，付红伟. 数据管道技术在系统数据备份与恢复中的应用. 软件导刊，2011，(4)：157～159.
[8] 王迪，舒继武，沈美明. 一种 SAN 环境下数据备份系统的设计与实现. 小型微型计算机系统，2006，(9)：1788～1792.
[9] 戴世剑，涂彦晖. 数据恢复技术. 2 版. 北京：电子工业出版社，2005.
[10] 胡敏，杨吉云，姜维. Windows 下基于文件特征的数据恢复算法. 计算机应用，2011，(2)：527～529.
[11] 文光斌. 数据恢复技术的发展前景、技术层次及常用方法. 网络安全技术与应用，2005(5)：74～76.
[12] 蔡立军. 计算机网络安全技术. 北京：中国水利水电出版社，2002.
[13] 梁亚声. 计算机网络安全教程. 北京：机械工业出版社，2008.
[14] 王改性，师鸣若. 数据存储备份与灾难恢复. 北京：电子工业出版社，2009.
[15] 李伟超，李焱，许利军. 计算机信息安全技术. 长沙：国防科技大学出版社，2010.

第7章　操作系统的安全

本章学习目标

随着计算机在日常生活中的普及，人们对计算机的依赖越来越强，计算机系统安全、数据安全成了必须引起高度重视的问题。通过备份数据、安装系统补丁、安装杀毒软件可以有效地保障计算机系统安全和数据安全。本章介绍计算机操作系统安全的基本概念，简单介绍安全操作系统的评价标准、单点登录的访问管理，并对当前操作系统安全技术进行全面介绍。

通过对本章的学习，应掌握以下内容：

(1) 了解操作系统安全的现状。

(2) 了解计算机安全等级及信息安全技术评估准则。

(3) 了解单点登录机制。

(4) 了解主流操作系统的主要安全机制。

7.1　操作系统安全性的基本概念

随着计算机技术与信息技术的发展，人们对计算机系统的依赖也愈来愈大。政府机关和企事业单位通常都将大量的重要信息高度集中地存储在计算机系统中。如何确保计算机系统中存储或传输数据的保密性、完整性以及系统的可用性早已成为信息系统亟待解决的重要问题，保障系统的安全性这一任务也落到了现代操作系统的肩上。值得注意的是，计算机网络虽然扩大了用户的通信范围和资源共享的程度，却增加了网络的复杂性和脆弱性，使网络更易受到别有用心者的攻击和破坏，所带来的损失也会更加严重。正因如此，系统安全性问题引起了国际上的广泛重视。近年来已开发出许多新的、可用于保障 Intranet 安全和在 Internet 上开展电子商务活动的安全协议和软件。通过备份数据、安装系统补丁、安装杀毒软件可以有效地保障计算机系统安全和数据安全。

7.1.1　操作系统的原理知识

操作系统的出现、使用和发展是近 40 余年来计算机软件技术的一个重大进步，它的出现为人们使用各种各样的计算机奠定了重要基础。计算机发展到今天，从个人机到巨型机，无一例外都配置一种或多种操作系统，操作系统已经成为现代计算机系统不可分割的重要组成部分，它为人们建立各种各样的计算机应用环境奠定了重要基础。计算机系统由硬件、软件和数据组成。在计算机系统的运行中，操作系统提供了合理利用这些资源的途径。可以从两个视角来研究操作系统：资源视角和用户视角。从资源管理的角度来看，操作系统是计算机系统中的资源管理器，负责对系统的软硬件资源实施有效的控制和管理，提高系统资源的利用率。从方便用户使用的角度来看，操作系统是一台虚拟机，是对计算机硬件的首

次扩充，隐藏了硬件操作细节，使用户与硬件细节隔离，从而方便用户使用。尽管操作系统尚未有一个严格的定义，但一般认为：操作系统是控制和管理计算机软硬件资源，以尽量合理有效的方法组织多个用户共享多种资源的程序集合。操作系统一般具有4个基本特征：并发性、共享性、虚拟性和不确定性。

1. 进程管理

不管是常驻程序或应用程序，它们都以进程为标准执行单位。最初运用冯·诺依曼体系结构建造计算机时，每个中央处理器最多只能同时执行一个进程。早期的OS（例如DOS）也不允许任何程序打破这个限制，并且DOS不能允许多个进程同时执行（虽然DOS自己宣称他们拥有终止并等待驻留（TSR）能力，可以部分且艰难地解决这问题）。然而，现代的操作系统即使只拥有一个CPU也可以利用多进程（Multiprocess）功能同时执行复数进程。进程管理指的是操作系统调整复数进程的功能。除了进程管理之外，OS还担负进程间通信（IPC）、进程异常终止处理、死锁（Dead lock）侦测及处理等任务。在进程之下还有线程的问题，但是大部分的OS并不会处理线程方面的问题，OS通常只会提供一组API让使用者自行操作或通过虚拟机器的管理机制控制线程之间的交互。进程状态演变如图7-1所示。

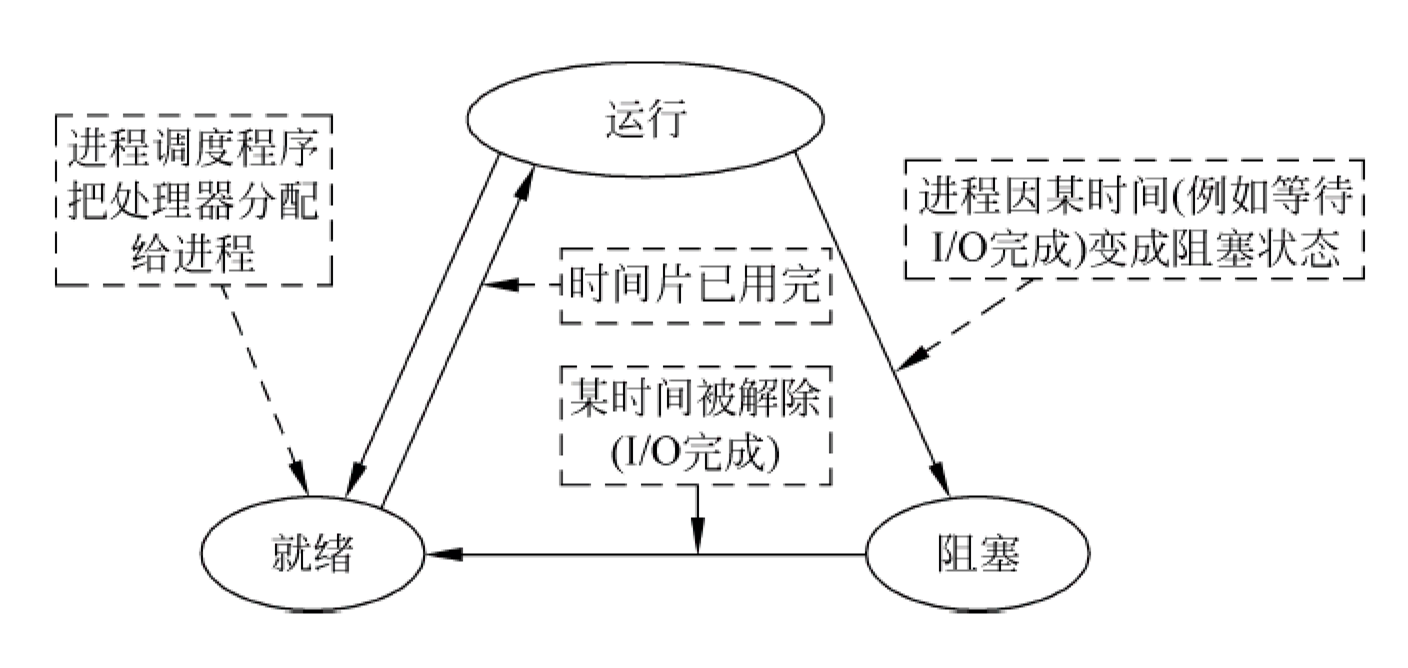


图7-1 进程状态演变图

2. 内存管理

大部分的现代计算机内存结构都是阶层式的，以最快且数量最少的寄存器为首，然后是高速缓存、内存以及最慢的磁盘存储设备。OS的内存管理提供寻找可用的记忆空间、配置与释放记忆空间、交换内存和低速存储设备的内含物等功能。这种又被称为虚拟内存管理的功能将大幅增加每个进程可获得的记忆空间（通常是4GB，即使实际上RAM的空间远少于4GB）。但这也带来了微幅降低执行效率的缺点，严重时甚至会导致进程崩溃。

3. 磁盘与文件系统

文件系统通常指的是管理磁盘数据的系统，其可将数据以目录或文件的形式存储。每个文件系统都有自己特殊的格式与功能。OS拥有多种内置文件系统，例如，Linux拥有非常广泛的内置文件系统，包括ext2、ext3、ReiserFS、Reiser4、GFS、GFS2、OCFS、OCFS2、NILFS与Google文件系统。Linux也支持非本地文件系统，包括XFS、JFS、FAT家族与NTFS。而Windows能支持的文件系统只有FAT12、FAT16、FAT32与NTFS。NTFS系统是Windows上最可靠和最有效率的文件系统，其他的FAT家族都比NTFS老旧，并且在文件长度与分割磁盘能力方面都有很大限制，因此造成很多问题。UNIX的文件系统多

半是 UFS,它的一个分支 Solaris 开始支持一种新式的 ZFS。

文件系统的操作原理如图 7-2 所示。

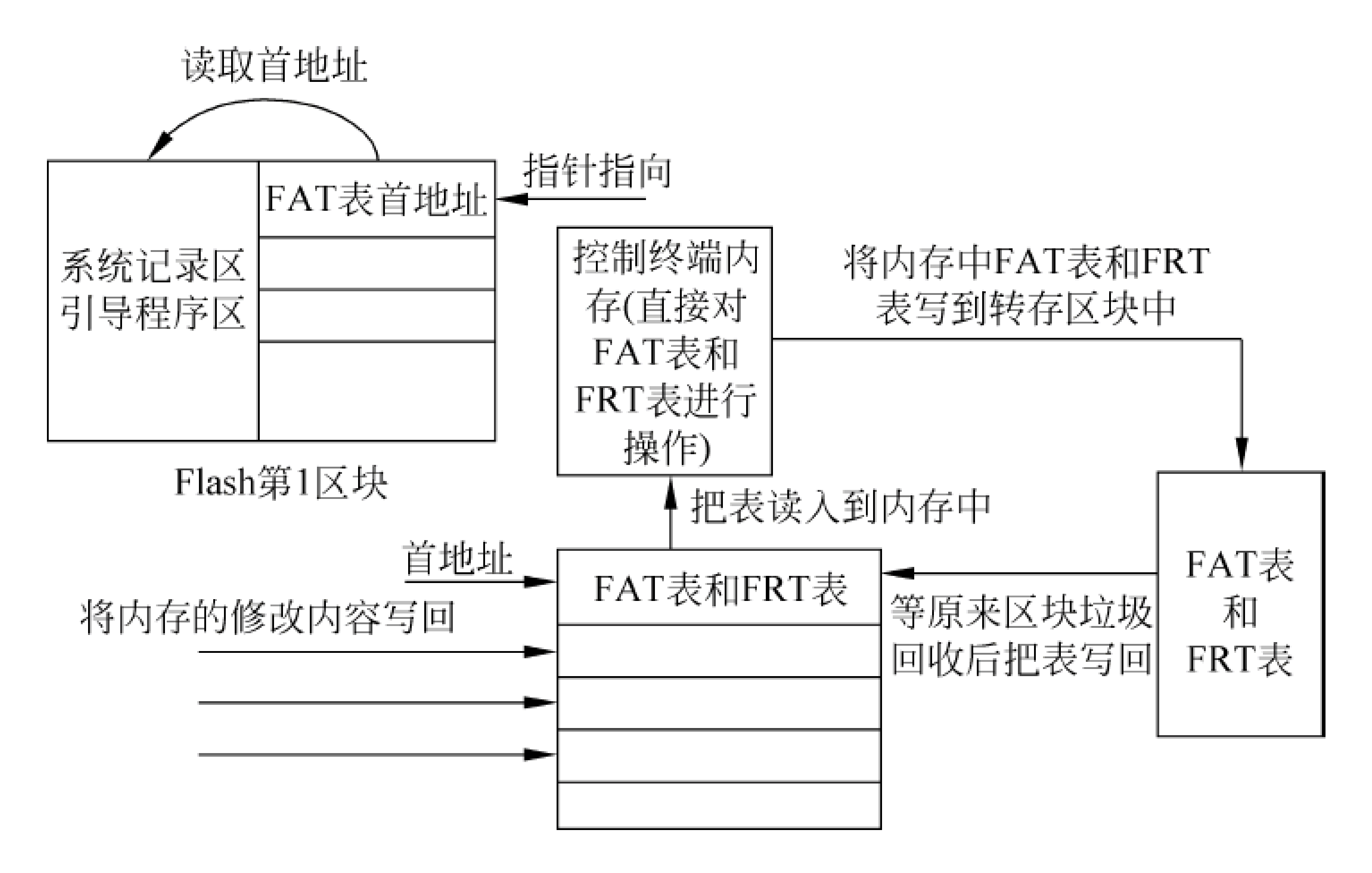


图 7-2　文件系统的操作原理框图

4. 网络

许多现代的 OS 都具备操作主流网络通信协议 TCP/IP 的能力,这就使得操作系统可以进入网络世界,并且与其他系统分享如文件、打印机与扫描仪等资源。许多 OS 也支持多个网络启蒙时代的各路网络通信协议,例如,IBM 建立的 SNA;DEC 在它所生产的系统所设置的 DECnet 结构;微软为 Windows 制作的特殊通信协议。还有许多为了特殊功能而研发的通信协议,例如可以在网络上提供文件存取功能的 NFS 系统、大量用于影音流及游戏消息传输的 UDP 协议等。

5. 安全

大多数 OS 都具有某种程度的信息安全机制。信息安全机制主要基于两大理念:OS 为外界提供直接或间接存取多种资源的管道,这些资源包括本地端磁盘机的文件、受保护的特权系统调用、使用者的隐私数据、系统执行的程序所提供的服务等;OS 有能力认证资源存取的请求,允许通过认证的请求并拒绝无法通过的非法请求。有些系统的认证机制仅简略地把资源分为特权或非特权,并且每个请求都有独特的身份辨识号码,例如使用者名称。资源请求通常分成两大类:内部来源与外部来源。内部来源通常是一个正在执行的程序发出的资源请求。在某些系统上,一个程序一旦可执行就可做任何事情(例如 DOS 时代的病毒),但 OS 通常会给程序一个识别代号,并且在此程序发出请求时,检查其代号与所需资源的存取权限关系。外部来源是从非本地端计算机传来的资源请求,例如远程登录本地计算机或某些网络连接请求(FTP 或 HTTP)。为了识别这些外部请求,系统也许会对此请求提出认证要求,通常是请求输入使用者名称以及相对应的密码。系统有时也会应用诸如磁卡或生物识别数据等多种认证方法。在某些情况下,例如网络通信时,通常不需要通过认证即可存取资源(例如匿名存取的 FTP 服务器或 P2P 服务)。除了允许/拒绝形式的安全机制,一个高安全等级的系统也会提供记录选项,允许记录各种请求对资源存取的行为。

6. 内部信息安全

内部信息安全可视为防止正在执行的程序任意存取系统资源的手段。大多数 OS 允许普通程序直接操作计算机的 CPU,因此会产生一些问题。例如怎样把类似 OS 一样处理事务、执行同样特殊指令的程序强迫停止。因为在这种情况下,OS 也只是另一个平起平坐的程序。为通用 OS 所生产的 CPU 通常在硬件层级上实践了一定程度的特殊指令保护概念。通常特权层级较低的程序想要执行某些特殊指令如直接存取硬盘之类的外部设备时会被阻断。因此,程序必须得经由 OS,让 OS 执行特殊指令来存取磁盘。从而也使得 OS 有机会检查该程序的识别身份,并依此接受或拒绝它的请求。

7. 外部信息安全

一个操作系统通常会为其他网络上的计算机或使用者提供各种服务。这些服务通常借由端口或 OS 网络地址后的数字接入点提供。通常此服务包括提供文件共享(NTFS)、打印共享、电子邮件、网页服务与文件传输协议(FTP)。外部信息安全的最前线是防火墙等的硬件设备。在 OS 内部也常常设置许多种类的软件防火墙。软件防火墙可设置接受或拒绝在 OS 上执行的服务与外界的连接。因此任何人都可以安装并执行某些不安全的网络服务,例如 Telnet 或 FTP,并且设置除了某些自用通道之外其他所有连接的阻挡,以防止不良连接。

7.1.2 安全操作系统评价标准

为了能有效地以工业化方式构造可信任的安全产品,国际标准化组织采纳了由美、英等国提出的"信息技术安全评价公共准则(CC)"作为国际标准。CC 为相互独立的机构对相应信息技术安全产品进行评价提供了可比性。

CC 由两部分组成,一部分是面向用户的信息技术产品的安全功能需求定义,用户可以按照安全功能需求来定义产品的保护框架(PP),CC 要求对 PP 进行评价以检查它是否能满足对安全的要求;CC 的另一部分是面向厂商的安全保证需求定义,厂商应根据 PP 文件制定产品的安全目标文件(ST),CC 同样要求对 ST 进行评价,然后根据产品规格和 ST 去开发产品。必须指出的是,保障计算机和系统的安全性将涉及许多方面,其中有工程问题、经济问题、技术问题、管理问题,甚至涉及国家的立法问题。

1. 可信任计算机系统评价标准(TCSEC)

对一个安全产品(系统)进行评估,是件十分复杂的事,它对公正性和一致性要求很严。因此需要有一个能被广泛接受的评估标准。美国国防部在 20 世纪 80 年代中期制定了一组计算机系统安全需求标准,共包括 20 多个文件,每个文件都使用了不同颜色的封面,统称为"彩虹系列"。其中核心的是具有橙色封皮的"可信任计算机系统评价标准(TCSEC)",简称为橙皮书。该标准将计算机系统的安全程度划分为 8 个等级:D1、C1、C2、B1、B2、B3、A1 和 A2。在橙皮书中,对每个评价级别的资源访问控制功能和访问的不可抵赖性、信任度及产品制造商应提供的文档作了一系列的规定:

(1) D 类安全等级。只包含 D1 类别,是最低的安全等级。D1 系统只为文件和用户提供安全保护。D1 系统最普通的形式是本地操作系统,或者是一个完全没有保护的网络。

(2) C 类安全等级。该类安全等级能够提供审慎的保护,并能为用户的行动和责任提

供审计能力。C 类安全等级又可划分为自主安全保护 C1 和可控访问保护 C2 两类。C1 系统的可信任计算基础(Trust Computing Base,TCB)将用户和数据分开以实现安全。在 C1 系统中,所有的用户以相同的灵敏度来处理数据,即所有用户认为 C1 系统中的所有文档都具有相同的机密性。相比之下,C2 系统进一步加强了可调的审慎控制。一旦连接到网络上,C2 系统的用户分别对各自的行为负责。通过登录过程、安全事件和资源隔离,C2 系统在具有 C1 系统中所有安全性特征的同时,控制力度进一步增强。

(3) B 类安全等级。B 类安全等级可分为标识安全保护 B1、结构安全保护 B2 和安全域保护 B3 三类。B 类系统具有强制性保护功能。强制性保护指的是如果用户没有与安全等级相连,系统就不会让用户存取对象。

(4) B1 系统。B1 系统需满足的要求有:系统对网络控制下的每个对象都进行灵敏度标记,系统使用灵敏度标记作为所有强迫访问控制的基础;系统对待导入系统中的非标记对象进行灵敏度标记,且其标记必须准确地表示所联系的对象的安全级别;当系统管理员创建系统或增加新的通信通道或 I/O 设备时,管理员必须手工制定每个通信通道和 I/O 设备是单级还是多级;系统必须使用用户的密码或证明来决定用户的安全访问级别;系统必须通过审计来记录未授权访问的企图。

(5) B2 系统。B2 系统必须先满足 B1 系统的所有要求。B2 系统的管理员还必须使用一个明确的、文档化的安全策略模式作为系统的可信任运算基础体制。此外,B2 系统必须立即通知系统中的每一个用户所有与之相关的网络连接的改变;只有用户能够在可信任通信路径中进行初始化通信;可信任运算基础体制能够支持独立的操作者和管理者。

(6) B3 系统。B3 系统必须符合 B2 系统的所有安全需求。B3 系统具有很强的监视委托管理访问能力和抗干扰能力。B3 系统必须设有安全管理员,同时需满足以下要求:除控制对个别对象的访问外,B3 必须产生一个可读的安全列表;每个被命名的对象提供对该对象没有访问权的用户列表说明;B3 系统在进行任何操作前,要求用户进行身份验证;B3 系统验证每个用户,同时还会发送一个取消访问的审计跟踪消息;系统的设计者必须正确区分可信任的通信路径和其他路径;可信任的通信基础体制为每一个被命名的对象建立安全审计追踪;可信任的运算基础体制支持独立的安全管理。

(7) A 类安全等级。A 系统是安全级别最高的系统。到目前为止,A 类安全等级只包含 A1 一个安全类别。A1 类与 B3 类相似,对系统的结构和策略不做特别要求。A1 系统的显著特征是系统的设计者必须按照一个正式的设计规范来分析系统。对系统分析后,设计者必须运用核对技术来确保系统符合设计规范。A1 系统必须满足以下要求:系统管理员必须从开发者处接收到一个安全策略的正式模型;所有的安装操作都必须由系统管理员进行;系统管理员进行的每一步安装操作都必须有正式文档。

TCSEC 在操作系统级上提出的可信计算机基础 TCB 包含的安全内容有:

(1) 操作系统内核。

(2) 具有特权的程序和命令。

(3) 具有处理敏感信息的程序。

(4) 与实施安全策略有关的文档资料。

(5) 保障硬件正确运行的程序和诊断程序。

(6) 构成系统的可信硬件。

(7) 负责管理系统的人员。

在安全操作系统设计过程中,上述内容是分别在各个层次中解决的。

2. 国内的安全操作系统评估标准

为了适应信息安全发展的要求,我国也制定了计算机信息系统登记划分准则:《信息技术安全性评估准则》GB/T 18336—2001。该准则将操作系统安全分为5个级别,分别是用户自主保护级、系统审计保护级、安全标记保护级、结构化保护级和访问验证保护级。5个级别对操作系统具备的安全功能有不同的要求,如表7-1所示。

表7-1 操作系统的5个安全级别

安全策略	用户自主保护级	系统审计保护级	安全标记保护级	结构化保护级	访问验证保护级
自主访问控制	√	√	√	√	√
身份鉴别	√	√	√	√	√
数据完整性	√	√	√	√	√
客体重用		√	√	√	√
审计		√	√	√	√
强制访问控制			√	√	√
标记			√	√	√
隐蔽信道分析				√	√
可信路径				√	√
可信恢复					√

上述5个级别从低到高,每个级别都实现上一级的所有功能,并且有所增加。具体内容可参考《信息技术安全性评估准则》GB/T 18336—2001,这里不再详述。

7.1.3 常见的系统安全保护方法

Internet的广泛应用使计算机病毒没有了国界,我国约有90%的计算机遭受过计算机病毒的攻击。面对计算机病毒的破坏,许多计算机用户束手无策。其实大可不必对病毒感到恐慌,只要对病毒有足够的认识并采取妥善的防治措施,计算机病毒就没有可乘之机,也就不会给用户造成无法挽回的损失了。下面就分别从数据备份、计算机病毒预防、病毒查杀、后期处理及防ARP攻击等方面介绍计算机系统安全的防护。

1. 备份数据

备份数据对用户的自有数据是非常重要的。用户的自有数据一旦遭到破坏,将造成不可弥补的损失,如果没有备份的话,后果不堪设想。因此建议各级用户都要及时妥善备份自有的数据,例如历年资料、重要方案、管理文献、重要数据等,并且要备份到本机之外的存储介质上(例如光盘、移动硬盘等)。

2. 病毒预防

(1) 大多数病毒都是通过操作系统的漏洞进入系统并产生危害的。

为了阻止病毒进入计算机系统,首先应该提高系统的自我保护能力,经常进行系统更新。执行"自动更新"时,Windows系统将例行检查Windows Update网站以获得高优先级更新,这些更新有助于保护计算机系统,防止它遭受最新病毒和其他安全威胁的攻击。这些

更新包括安全更新、重要更新和 Service Pack。Windows 会自动下载并安装计算机所需要的所有高优先级更新。这些操作可设置成自动执行。给系统打补丁能减少黑客和病毒进入系统的可能性。但这并非就可以不用采取任何其他措施防范病毒了。由于发现系统漏洞需要一个时间过程,因此还需采取其他手段预防病毒。

(2) 病毒通过与其他计算机进行数据交换而进入系统。

预防病毒最好的方法是断开本机与其他计算机的任何数据交换,但这在实际使用过程中是不可能的。大量与外界交换信息,就给病毒的传播与感染创造了条件。如果用户的工作必须要与外界进行数据交换,或者是需要从外界获得大量数据,那么安装防杀病毒的软件是目前最普遍的保护方式。当前国内外众多杀毒软件开发商(例如江民、瑞星、金山、卡巴斯基、Vast 等)都开发了功能完善、操作简单的杀毒软件,用户可以通过在机器上正确安装病毒防火墙和查、杀病毒软件,开启杀毒软件的实时监控功能来做好预防病毒的工作。目前,绝大部分杀毒软件实时监控功能安装后默认自动开启。在计算机的使用过程中,不要轻易使用来历不明的各种软件;不要打开、运行来历不明的 E-mail 附件,尤其是在邮件主题中以诱惑的文字建议用户执行的邮件附件程序;及时升级杀毒软件,确保所使用的查、杀病毒软件的扫描引擎和病毒代码库为最新的,定期使用杀毒软件扫描系统。有了杀毒软件,并不表示用户就可以万事无忧了,因为新病毒是不断出现的,系统仍有可能感染病毒。

3. 病毒查杀

在一般情况下,计算机病毒总是依附某一系统软件或用户程序进行繁殖和扩散的,病毒发作时危及计算机的正常工作,破坏数据与程序,侵犯计算机资源。计算机在感染病毒后,总是有一定规律地出现异常现象,例如,屏幕显示异常,屏幕显示的不是由正常程序产生的画面或字符串;读写磁盘时间比平时增长,磁盘出现莫名其妙的文件和坏块,内存空间减小;系统运行速度下降,丢失数据或程序,文件字节数发生变化;系统自行引导,异常死机;系统引导时间增长,用户没有访问的设备出现工作信号等。如果出现上述现象时,系统就很有可能被病毒感染了,这时应当停止对计算机的任何操作,启动杀毒软件,对整个硬盘进行病毒查杀。部分计算机病毒(引导型病毒)在 Windows 系统启动状态下无法完全被清除,此时应使用事先准备好的杀毒软件应急盘完成杀毒。杀毒完成后,重启计算机,再次启动杀毒软件检查系统是否还存在病毒,并确定被感染破坏的数据确实被完全恢复。特殊情况下还需要断开网络,在安全模式下杀毒才能将病毒清理干净。

4. 后期处理

虽然杀毒软件的功能强大,但经过杀毒软件处理的染毒文件有时可能会导致系统不能正常运行,有些虽然可以运行,但其稳定性会降低,程序崩溃的机会增加,更有甚者可能造成重要的自有数据的破坏或丢失。如果是一般性质的病毒,利用杀毒软件可以清除,对于系统和数据影响不会很大;如果受破坏的是系统软件,并且染毒程度比较重,可能导致系统不能启动或正常使用。在这种情况下,可以先杀毒,然后针对不同的操作系统采取不同的措施:对于 Windows 98,可以进行修复安装;对于 Windows 2000/XP,可以在“开始”|“运行”中输入“sfc/scannow”来对系统文件进行修复。如果问题得不到解决,还可以对操作系统进行修复性安装,一般情况下均可解决问题。需要注意的是,如果感染的是重要的数据文件,且破

坏得较为严重,应该请专业人士进行数据挽救。

5. 防 ARP 攻击

ARP 攻击就是通过伪造 IP 地址和 MAC 地址实现 ARP 欺骗,从而在网络中产生大量的 ARP 通信流量使网络阻塞。攻击者只要持续不断地发出伪造的 ARP 响应包就能更改目标主机 ARP 缓存中的 IP-MAC 条目,造成网络中断或中间人攻击。ARP 欺骗木马的中毒现象表现为:使用局域网时会突然掉线,过一段时间后又会恢复正常;客户端状态频频变红,用户频繁断网,IE 浏览器频繁出错,以及一些常用软件出现故障等;如果在局域网中是通过身份认证上网的,会突然出现可认证但不能上网的现象,重启机器或在 MS-DOS 窗口下运行命令 arp -d 后,又可恢复上网。

ARP 欺骗木马只需成功感染一台计算机,就可能导致整个局域网都无法上网,严重的甚至可能带来整个网络的瘫痪。该类木马发作时除了会导致同一局域网内的其他用户上网出现时断时续的现象外,还会窃取用户密码。木马的惯用伎俩有盗取 QQ 密码、盗取各种网络游戏密码和账户去做金钱交易、盗窃网上银行账户去做非法交易活动等,给用户造成了很大的不便和巨大的经济损失。ARP 攻击主要存在于局域网网络中,局域网中若有一个系统感染 ARP 木马,该系统将会试图通过 ARP 欺骗手段截获所在网络内其他计算机的通信信息,并因此造成网内其他计算机的通信故障。目前关于 ARP 类的防护软件有多种,比较常用的 ARP 工具主要用来检测 ARP 攻击,其工作原理是以一定频率向网络广播正确的 ARP 信息。

用来保障计算机和系统安全的基本技术包括认证技术、访问控制技术、密码技术、数字签名技术和防火墙技术等。但在本章以下内容中我们主要介绍认证技术、访问控制技术和防火墙技术。

7.2 单点登录的访问管理

1. 单点登录的概念

单点登录(Single Sing On,SSO)是目前比较流行的企业业务整合的解决方案之一,指的是在多个应用系统中,用户只需要登录一次就可以访问所有相互信任的应用系统。

较大的企业内部一般都有很多业务支持系统为其提供相应的管理和 IT 服务,如财务系统、人事系统等。这些系统往往是在不同时间建设起来的,运行在不同的平台上,还可能使用了不同的技术和标准。随着企业的发展,业务系统的数量在不断增加,需要维护的系统越来越多,管理上的开销越来越大。为了降低管理的消耗,最大限度地重用已有的系统,很多企业都在进行着企业应用集成(EAI)。企业应用集成可以在不同层面上进行,如数据存储层面上的"数据大集中",传输层面上的"通用数据交换平台",在应用层面上的"业务流程整合"和用户层面上的"通用企业门户"等。事实上,"身份认证"集成也变得越来越重要,这也就是实现单点登录的目的。

通常情况下,每个单独的系统都有自己的安全体系和身份认证系统。在集成之前,进入每个系统都需要进行登录,从而给管理带来了很大困难,在安全方面也埋下了重大隐患。使用单点登录后,只需要登录一次就可以进入多个系统,而不需要重新登录,这不仅带给用户

更好的体验,同时也降低了安全风险和管理消耗。

根据登录的应用类型不同,可将单点登录分为两种类型。

1）对桌面资源的统一访问管理

包括两个方面：

(1）登录 Windows 后统一访问 Microsoft 应用资源。

(2）登录 Windows 后访问其他应用资源。

2）Web 单点登录(Web-SSO)

由于 Web 技术体系结构便捷,对 Web 资源的统一访问管理易于实现,如图 7-3 所示。

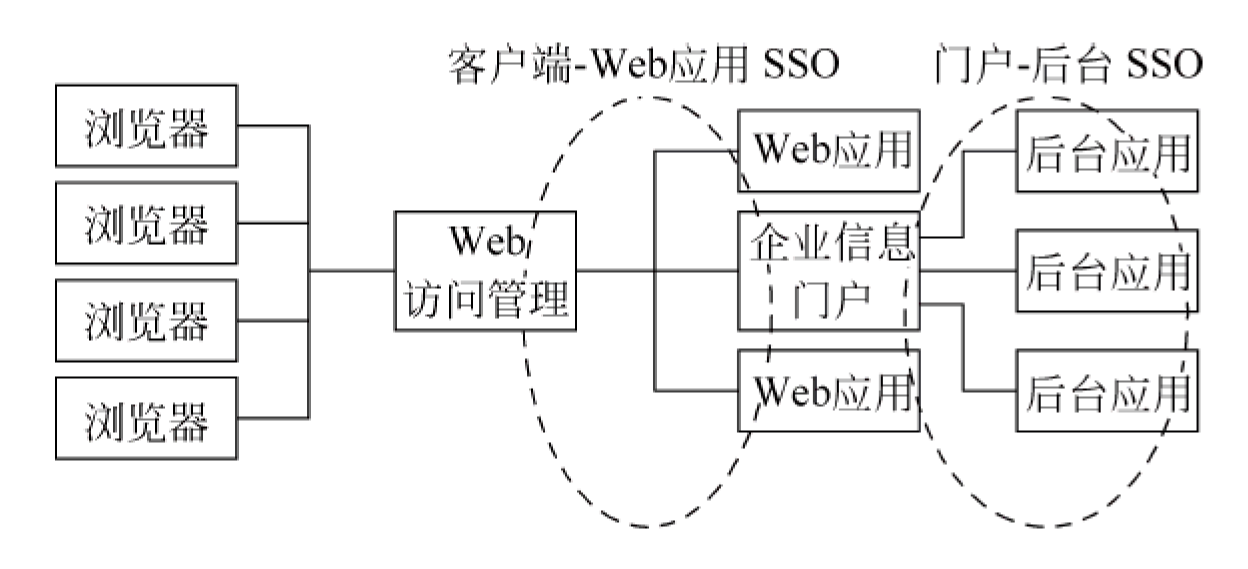


图 7-3　Web 单点登录访问管理系统

2. SSO 实现机制

第一次访问应用系统 A 时,由于还没有登录,用户会被引导至认证系统中进行登录。根据用户提供的登录信息,认证系统进行身份验证,若通过则认证系统返回给用户一个认证的凭据(Ticket)。用户再访问别的应用时,会带上这个 Ticket 作为自己认证的凭据。应用系统接收到请求之后将 Ticket 送入认证系统进行验证,若合法则通过验证,用户就可以在不用再次登录的情况下访问系统 B 或系统 C 了。

要实现 SSO,需要实现以下主要功能：

1）所有应用系统共享一个身份认证系统

统一的认证系统是 SSO 的一个前提。认证系统的主要功能是将用户的登录信息和用户信息库进行比对,从而对用户进行登录认证。认证成功后,认证系统应该生成统一的认证标志(Ticket)返回给用户。同时,认证系统还需对 Ticket 进行验证,判断其有效性。

2）所有应用系统能够识别和提取 Ticket 信息

要实现 SSO 的功能,让用户只需登录一次,就必须使应用系统能够识别已经登录过的用户。应用系统需要对 Ticket 进行识别和提取,通过与认证系统的通信,自动判断出当前用户是否已经登录过,从而完成单点登录的功能。

需要注意的是,单一的用户信息系统并不是必须具备的。许多系统不能将所有的用户信息都集中存储,应当允许用户信息存放在不同的存储中。此外,统一的认证系统并不仅限于单个认证服务器。具有多台认证服务器时,服务器之间要通过标准的通信协议,相互交换认证信息,从而实现更高级别的单点登录。

3. Web-SSO 的实现

用户在访问页面 A 时进行了登录,但是客户端的每个请求都是单独的连接,当用户再次访问页面 B 的时候,如何才能通知 Web 服务器客户刚才已经登录过了呢？由于浏览器和

服务器之间通过使用Cookie技术来维护应用的状态,Cookie是可以被Web服务器设置的字符串,并且可以保存在浏览器中。当浏览器访问了页面A时,Web服务器设置了一个Cookie,并将这个Cookie和页面A一起返回给浏览器,浏览器保存返回的Cookie,在访问页面B时带上这个Cookie,Web服务器接到请求时也能读出Cookie的值,根据Cookie值的内容判断和恢复一些用户的信息状态。Web-SSO完全可以利用Cookie技术来完成用户登录信息的保存,将浏览器中的Cookie和Ticket结合起来,完成SSO的功能。

为了完成一个简单的Web-SSO的功能,需要以下两个部分的合作:

(1) 统一的身份认证服务。

(2) 修改Web应用,使得每个应用都通过这个统一的认证服务来进行身份校验。

实现Web-SSO的技术主要有:

(1) 基于Cookie实现。该方法可以基于数据库实现,可以实现跨域的SSO。但是,基于两个域名之间传递sessionID的方法可以在Windows中成立,在UNIX和Linux中可能出现问题,同时在某些应用下安全性较差。

(2) Broker-based(基于经纪人),如Kerberos、Sesame、IBM KryptoKnight等。其特点是具有一个集中的认证和用户账户管理的服务器。Kerberos是由麻省理工大学发明的安全认证服务,已经被UNIX和Windows作为默认的安全认证服务集成进入操作系统。

(3) Agent-based(基于代理人),如SSH等。该方案具有一个自动为不同的应用程序认证用户身份的代理程序。这个代理程序需要设计不同的功能,如可以使用密码表或加密密钥自动地将认证的负担从用户方移开。代理人被放置在服务器上,在服务器的认证系统和客户端认证方法之间充当一个"翻译"。

(4) Token-based(基于票据),如SecurID、WebID等。该方法为当前广泛使用的密码认证,如FTP、邮件服务器的登录认证等,是一种简单易用的方式,实现一个密码在多种应用中使用。

(5) 基于网关。

(6) 基于安全断言标记语言(SAML)。SAML的出现大大简化了SSO,并被结构化信息标准促进组织(OASIS)批准作为SSO的执行标准。

7.3 主流操作系统的安全性

安全操作系统的具体实施与该系统的设计目标和应用目标有关,也与其运行的平台有关。并非所有的应用都需要A1级的安全性,实际上也很少有操作系统具备能使应用达到A1级的特性。在实际的实施中,考虑重点是广泛使用的系统的安全内核以及存取控制。接下来主要介绍目前使用最广泛的两类操作系统的安全特性。

7.3.1 UNIX/Linux的安全

1. UNIX安全

UNIX由Ken Thompson、Dennis Ritchie和Douglas Mcllroy于1969年在AT&T的贝尔实验室开发。UNIX是一个强大的多用户、多任务操作系统,支持多种处理器结构。

UNIX 的开发者在开发初期并不打算使该系统拥有很高的安全性，因为其应用基于相互信任的环境，例如研究所、实验室、大学等。在这些场所，共享带来的便利远远大于不友好访问的威胁。

因此，系统中采用了一般的安全机制，文件、数据、设备、存储卷的共享相对简单，不采用强保护机制。UNIX 的超级权限是“超级用户”，超级用户能完成系统中的任何操作，因此也成为攻击的对象。攻击者一旦获得超级用户权限，就能建立在以后任何时间提供超级用户存取的密钥。UNIX 系统存在的隐蔽通道暴露出了自身的一些漏洞。Morris 蠕虫正是利用了这些安全缺陷给数以万计的用户造成了巨大损失。用于特殊环境安全的 UNIX 操作系统实际是通过重写 UNIX 内核，并提供具有不同内部结构的 UNIX 外部功能实现的。

目前，UNIX 系统的安全性在不断增加，并出现了许多安全检测工具，例如 Quest、UXA、Alert/Inform、Sfind、USECURE、Kerberos 等。系统管理员通过安全检测工具检测安全机制、权限和安全域设置、可疑入侵和特洛伊木马等。在目前的 UNIX 系统中，常规 UNIX 具有 C1 级安全级别，OSF/1 具有 B1 的安全级别，USL 的 SVR4/ES 则具有 B2 的安全级别。

2. Linux 安全

Linux 系统由芬兰赫尔辛基大学的学生 Linux Torvalds 于 1991 年开发。Linux 类似 UNIX 系统，并具有源码开放、免费使用和可自由传播等特点。经过 20 多年的发展，Linux 的安全机制日益完善，按照 TCSEC 的评估标准，目前 Linux 的安全级基本达到了 C2 级。Linux 的安全机制主要有 PAM 机制、文件系统加密、入侵检测机制、安全日志文件机制、强制访问控制和防火墙机制等。

1）PAM 机制

PAM(Pluggable Authentication Modules)是一套共享库，提供一个框架和一套编程接口给系统管理员，由系统管理员在多种认证方法中选择适宜的认证方法，并能够改变本地认证方法而无须重新编译与认证相关的程序。PAM 的主要功能有：密码加密，根据需要限制用户对系统资源的使用，防止拒绝服务攻击；支持 Shadow 密码；限制特定用户在指定时间从指定地点登录；引入概念“client plug-in agents”，使 PAM 支持 C/S 应用中的“机器-机器”认证成为可能；PAM 为更有效的认证方法的开发提供了便利，在此基础上可以很容易地开发出替代常规的用户名加密码的认证方法，例如智能卡、指纹识别等认证方法。

2）文件系统加密

加密技术在现代计算机系统安全中扮演着越来越重要的角色。文件系统加密是将加密技术应用到文件系统，从而提高计算机系统的安全性。目前的 Linux 已具有多种加密文件系统，例如 CFS、TCFS、CRYPTFS 等，其中具有代表性的是 TCFS(Transparent Cryptographic File System)。TCFS 将加密服务和文件系统相集成，用户在使用过程中并不会感觉出文件的加密过程。此外，TCFS 不修改文件系统的数据结构，备份与修复及用户访问保密文件的语义也不发生改变。TCFS 能使合法拥有者以外的用户、用户和远程文件系统通信线路上的窃听者以及文件系统服务器的超级用户不可读取其保密文件。而对于合法用户，访问保密文件与访问普通文件几乎没有区别。

3）入侵检测机制

Linux 的较新版本配备了入侵检测工具，其中比较流行的入侵检测系统有 Snort、Portsentry 和 Lids 等。所具有的入侵检测能力包括：记录入侵企图，当攻击发生时及时通知管理员；当预先定义的攻击行为发生时，采取预定义的措施处理；发出一些错误信息，例如伪装成其他操作系统，误导攻击者使得他们以为正在攻击一个 Windows NT 或 Solaris 系统，以增加攻击的难度。

4）安全日志文件机制

由于在各种安全措施检测下的操作系统仍然会有新漏洞出现。攻击者在漏洞被修补之前会乘机攻破尽可能多的机器。由于 Linux 不能预测主机遭受攻击的时机，但是可以记录攻击者的行踪。这些信息可以在日志中查询到。

日志是 Linux 安全结构中的一个重要内容，它是提供攻击发生的唯一真实证据。由于当前的攻击方式多种多样，因此 Linux 会记录网络、主机和用户级的日志信息，包括所有系统和内核信息、每一次网络连接和它们的源 IP 地址及长度、远程用户申请访问的文件、用户可以控制的进程、具体用户使用的每条命令等。日志信息是调查网络入侵者时不可缺少的证据，显然这属于一种事后调查。

5）强制访问控制

由于 Linux 是一种自由操作系统，当前在其平台上实现强制访问控制的产品有 SELinux、RSBAC、MAC 等，采用的策略也各不相同。

NSA 推出的 SELinux 安全体系结构称为 Flask，安全性策略的逻辑和通用接口一起封装在与操作系统独立的安全服务器中。SELinux 的安全服务器定义了一种混合的安全性策略，由类型实施（TE）、基于角色的访问控制（RBAC）和多级安全（MLS）组成。通过替换安全服务器，可以支持不同的安全策略。SELinux 使用策略配置语言定义安全策略，然后通过 checkpolicy 编译成二进制形式，存储在文件/ss_policy 中，在内核引导时读到内核空间。这就意味着安全性策略在每次系统引导时都会有所不同。

RSBAC（Rule Set Based Access Control，基于规则集的访问控制）是根据 Abrams 和 LaPadula 提出的 Generalized Framework for Access Control（GFAC）模型开发的，可以基于多个模块提供灵活的访问控制。所有与安全相关的系统调用都扩展了安全实施代码，这些代码调用中央决策部件，再由该部件调用所有激活的决策模块，形成一个综合的决定，最后由系统调用扩展来实施这个决定。RSBAC 目前包含的模块主要有 MAC、RBAC、ACL 等。MAC 是英国的 Malcolm Beattie 针对 Linux 2.2 编写的一个初级的 MAC 访问控制，将一个运行的 Linux 系统分割成多个互不可见或互相限制的子系统，这些子系统可以分别看作单一的系统来进行管理。

6）防火墙机制

防火墙是在被保护网络和 Internet 之间，或在其他网络之间限制访问的一种重要部件。Linux 防火墙系统提供了访问控制、审计、抗攻击和其他附属功能。其中，实现访问控制的方法是执行基于地址（源和目标）、用户和事件的访问控制策略，从而可以禁止非授权的访问，同时还能保护内部用户的合法访问。对通过该防火墙的网络访问进行记录，建立完备的日志、审计和追踪网络访问记录，产生报表等以实现审计功能。由于防火墙系统直接暴露在非信任网络中，对外界来说所有的攻击都是针对受到防火墙保护的内部网络这一个点。因

此要求被称为堡垒机的这一个点具有高度的安全性和抗攻击能力。除了上述功能外，与审计相关的报警和入侵检测，与访问控制相关的身份认证和加密等附属功能也需要防火墙来实现。

7.3.2　Windows 2000/XP 的安全

Windows NT 是美国 Microsoft 公司于 1992 年发布的 32 位操作系统，其安全级别达到 TCSEC 的 C2 级。Windows NT 的安全模型由本地认证、安全账户管理和安全参考监视器以及注册、访问控制和对象安全服务等构成。作为 Windows NT 的后续版本，Windows 2000/XP 提供了更多的新的安全机制。

1. 活动目录（Active Directory，AD）服务

活动目录服务在网络安全中具有极其重要的作用。活动目录为用户、硬件、应用以及网络上传输的数据提供了一个存储中心。活动目录还存储用户的授权和认证信息。Windows 2000 活动目录存储信息采用的是一种逻辑分层结构，如图 7-4 所示，该结构具有很好的扩展性和简化管理。活动目录使用域（Domains）、组织单元（Organizational Units，OU）和对象组织网络资源，这与 Windows 用文件夹和文件组织 PC 本地信息类似。

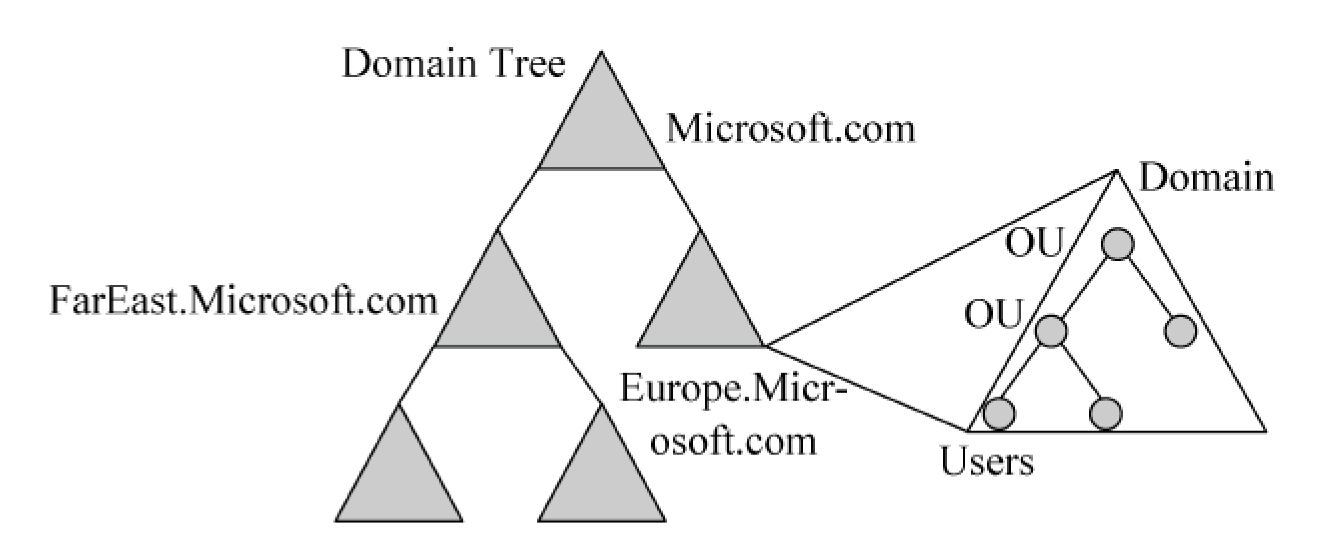


图 7-4　活动目录服务的分层结构

一个域是一个网络对象，是组织单元、用户账户、组合计算机的集合，多个域组成域树（Domain Tree）。组织单元是将对象组织成逻辑管理组的容器，可包含一个或多个对象。一个对象是一个独立的个体，这些个体可以是用户账户、计算机、扫描仪等。Windows 2000 以域间的信任关系控制用户对网络资源的访问权限。通过建立域间的信任关系，可以允许用户和计算机在任何一个域中进行身份认证从而使用经过授权的资源，这种横穿多个域保持的信任关系又称为穿越信任。采用穿越信任关系可以大大地减少网络中单项信任关系的量，从而简化网络管理。

2. Kerberos 审计协议

Windows 2000 使用 Internet 标准 Kerberos V5 协议（RFC 1510）进行用户审计。Kerberos 协议定义了客户端和密钥分配中心的认证服务之间的安全交互。Kerberos 协议为客户机/服务器建立连接前提供一种交互审计的机制，其特点有以下几点：

(1) 在建立初始连接时增强服务器认证性能。应用服务器不需要连接到域控制认证客户端，这样应用服务器可处理大量的客户连接请求，同时具有良好的可伸缩性。

(2) 多层客户机/服务器应用的认证委派。

(3) 具有穿越信任关系的域间认证。

Kerberos 认证基于票据(Tickets),当某客户登录到基于 Window 2000 的域中时获得一个票据,用于验证客户正在访问的网络资源的合法性以及客户能访问的网络资源。Kerberos 协议依赖于共享密钥的认证机制,客户和服务器都需要注册到 Kerberos 的认证服务器上。基本过程为:客户登录并从 KDC(Key Distribution Center,密钥分配中心)得到一个许可票据(Granting Ticket);客户用许可票据在 KDC 处为每种网络资源访问的会话获取会话票据(Session Tickets);客户在与应用服务器建立连接时提交会话票据;应用服务器验证 KDC 签发的会话票据的合法性以确定是否建立客户机/服务器连接。

3. PKI(Public Key Infrastructure,公钥基础设施)

公钥加密主要用在 Internet 一类的开放网络,用户通过证书进行数据加密、数字签名和身份验证。公钥加密技术的挑战在于跟踪证书。PKI 提供使用、管理和发布公钥证书的服务。Windows 2000 PKI 为用户提供基于 PKI 技术的超强安全系统,支持公钥加密服务。用户并不需要知道证书究竟是如何存储和工作的,证书完全与活动目录及操作系统分布式安全服务集成,其基本组件包括以下几个部分:

(1) 证书服务。允许组织和企业建立自己的 CA 系统,发布和管理数字证书。

(2) 活动目录。提供查找网络资源的唯一位置,提供证书和信息发布的服务。

(3) 基于 PKI 的应用。包括 Internet Explorer、Internet Information Server 和 Outlook Express 等。还有一些第三方 PKI 应用。

(4) 交换密钥管理服务(Exchange Key Management Service)。允许应用存储和获取用于加密 E-mail 的密钥。

Windows 2000 PKI 提供的安全功能具有互操作性、安全性、灵活性以及易用性等特点。

4. 智能卡

智能卡是用一种相对简单的方式使非授权人更难获得访问网络的权限。采用智能卡进行认证时,用户把卡插入连接到计算机的读写器中,并输入卡的 PIN(Personal Identification Number,个人标识号),Windows 使用卡中存储的私钥和证书向 Windows 域控制器的 KDC 认证用户,认证完毕后 KDC 返回许可票据。智能卡认证比密码认证具有更高的安全性,其原因有以下几点:

(1) 智能卡需要一个物理卡来认证用户。

(2) 智能卡的使用必须提供一个个人标识号来保证只有合法授权用户使用该智能卡。

(3) 由于无法从智能卡中提取出密钥,因此可以防范攻击者盗用用户证书。

(4) 能防止攻击者访问智能卡保护的重要资源。

(5) 没有密码或任何可重用信息的传输。

5. 加密文件系统(Encrypting File System,EFS)

Windows 的加密文件系统支持用户对指定的本地计算机中的文件或文件夹进行加密,防止非法用户对加密文件的读写操作。此外,当计算机物理丢失时,EFS 系统也可防止敏感信息的丢失和泄露。

Windows 使用 CryptoAPI 提供的公钥和对称密钥加密算法对文件或文件夹进行加密。用户使用 EFS 时并不需要了解其复杂的内部实现机制,只需要选中文件或文件夹后,在属性菜单中选中相应的菜单项即可完成加密。当合法用户再次打开文件时就会自动解密,而

非授权用户就无法读写加密文件。

6. 安全配置模板(Security Configuration Templates)

Windows 提供安全模板工具(Security Templates Tool)来组织网络的建立和管理。系统管理员使用管理控制台定义标准模板并统一地应用到多个计算机和用户中。

一个安全模板是一个安全配置的物理表示,即是存储一组安全设置的文件。Windows 包含了一组标准安全模板,从低安全的客户端配置至高安全的域控制器配置,用于不同场合和不同角色的计算机。这些模板可被直接使用或作为用户定制安全模板的基础。一个模板中一般应包括以下安全设置项:

(1) 账户策略。包括账户和密码锁定以及 Kerberos 策略的安全性。

(2) 本地策略。用户权限和记录安全事件。

(3) 事件日志。定义事件日志的安全性。

(4) 受限组。本地组成员的管理。

(5) 注册表。本地注册表项的安全性。

(6) 文件系统。本地文件系统的安全性。

(7) 系统服务。本地服务的安全性和启动模式。

思　考　题

(1) 计算机操作系统的基本原理知识有哪些? 掌握这些知识对于用户了解操作系统安全有何帮助?

(2) 当今的操作系统面临哪些安全威胁? 如何对这些安全威胁进行防御?

(3) 操作系统的安全有哪些积极作用? 面对当今的网络高速发展现状,如何确保操作系统的安全? 给出相应的分析。

(4) 单点登录的应用类型有哪两种?

(5) Web-SSO 的实现机制是什么? 有哪些关键技术?

(6) UNIX 的安全级别是什么? 主要的安全机制有哪些?

(7) Linux 的安全级别是什么? 主要的安全机制有哪些?

(8) Windows 2000/XP 的安全机制有哪些?

(9) 智能卡的安全性表现在哪些方面?

参 考 文 献

[1] 莫瑞加瑟著. 计算机安全的技术与方法. 吴亚非,等译. 北京: 电子工业出版社,1992.

[2] 王宁. 信息保障的理解和实践. 信息网络安全,2004,(1): 22～25.

[3] 李晓东,阎保平. 计算机网络信息管理及其安全. 微电子学与计算机,2002,19(5): 31～33.

[4] 焦俊梅. 浅淡网络信息的安全措施. 中山大学学报,2003,(5): 82～83.

[5] G Denker,J Millen,Y Miyake. PKI and Revocation Survey. Computer Science Laboratory,2000.

[6] 季庆光，唐柳英.结构化保护级安全操作系统安全策略模型. 北京：中科院信息安全技术工程研究中心，中软网络技术股份有限公司，2002.
[7] 卿斯汉.操作系统安全导论.北京：科学出版社，2003.
[8] 石文昌.安全操作系统研究的发展.计算机科学.2002，(7)：9～12.
[9] 贾铁军.网络安全实用技术.北京：清华大学出版社，2011.
[10] 郝玉洁，刘贵松，秦科，晏华.信息安全概论.成都：电子科技大学出版社，2007.

第 8 章　Web 站点的安全

本章学习目标

随着经济全球化的发展,信息成为商家、政府、乃至个人之间相互竞争的重要因素,而 Web 站点是信息交换和发布的重要工具。因此,如何保障 Web 站点的安全已成为目前关注的焦点。本章主要介绍网络安全的概念、机制、体系及技术,详细解释 Web 应用程序上的安全问题及漏洞,介绍基于 IIS 和 ASP 网站安全体系的建立及技术实现,最后讨论防火墙技术在站点的应用。

通过对本章的学习,应掌握以下内容:

(1) 了解 Web 网络安全的概念、机制和特点。

(2) 认识 Web 站点的安全隐患。

(3) 掌握网络安全的一些安全漏洞及测试。

(4) 描述有利于 Web 站点安全的技术实现,例如 IIS 和 ASP 网站安全体系的建立及技术实现。

(5) 了解防火墙在 Web 站点安全中的应用。

Internet 为信息共享、信息交流、信息服务创造了理想的空间,同时也带来了一定的风险。网络安全性由一系列不断发展的技术组成,特别是应用在通信或者商务服务的 Internet 相连接的网络中。对于 Web 系统开发和管理人员来说,保证 Web 站点的安全是一个关键而又复杂的问题。

本章将从 Web 站点的信息系统安全的概念入手,介绍它的机制和某些体系,分析当前 Web 站点所面临的安全问题,对常见的网络安全技术及有效的安全漏洞检测方法进行描述,介绍当前比较常用的 IIS 和 ASP 网站安全的建立和维护方法。

8.1　Web 的基本概念

8.1.1　Internet

Internet 起源于冷战时期美国国防部的高级研究计划局 ARPA(Advanced Research Project Agency)的试验网络 ARPANET。该网站建立于 1969 年,由 4 台计算机互联而成。到 1976 年,该网的节点机已发展到 57 个,连接不同类型的计算机 100 多台。ARPANET 最初采用“主机-主机”协议,后改为“网络控制协议”NCP。1982 年 ARPANET 、MILNET 等几个计算机网络实现互联,并决定采用网络互联协议 IP(Internet Protocol),并由此称为 Internet。

Internet 提供的服务可归纳为以下 9 项:

(1) 远程登录服务。远程登录是 Internet 最早提供的基本服务功能之一。Internet 中的用户远程登录是指使用 Telnet 命令,使自己的计算机暂时称为远程计算机的一个仿真终

端的过程。用户必须是远程终端的合法用户,并拥有相应的账户和密码。Telnet 协议是 TCP/IP 协议的一部分,它详细定义了客户机与远程服务器之间的交互过程。它的主要优点是能够解决不同类型的计算机系统之间的互操作问题。Telnet 协议引入了网络虚拟终端(Network Virtual Terminal,NVT)的概念,提供了一种专门的键盘定义,用来屏蔽不同计算机系统对键盘输入的差异性。Telnet 采用了客户机/服务器模式,其结构如图 8-1 所示。

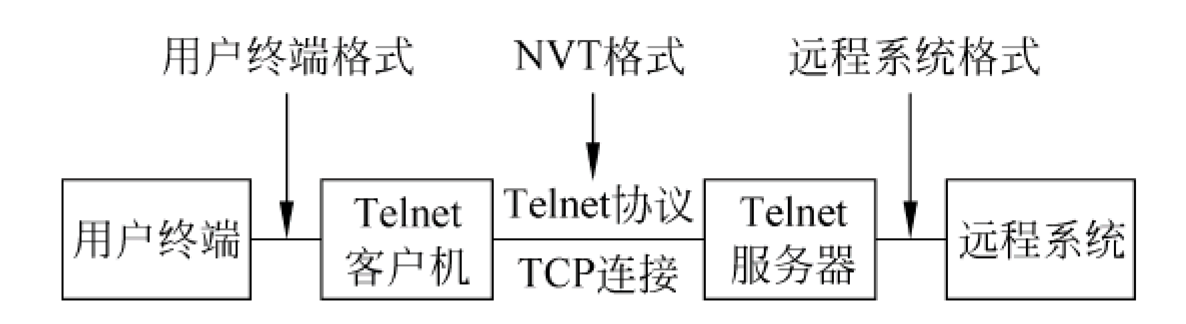


图 8-1 Telnet 协议客户机/服务器模式结构图

(2) 文件传输服务 FTP。FTP 服务由 TCP/IP 的文件传输协议(File Transfer Protocol)支持。只要加入 Internet 网的两台计算机都支持 TCP/IP 协议,无论它们相距多远,用户都能将一台计算机上的文件传输到另一台计算机上。普通的 FTP 服务要求用户在登录到远程计算机时提供相应的用户名和密码。也有一些 FTP 服务器无须用户事先申请用户名和密码,容许用户以 anonymous 作为用户名,用自己的 E-mail 地址作为密码,这种 FTP 服务称为匿名服务。图 8-2 所示为 FTP 在本地和远程文件系统之间传输文件的简单示意图。

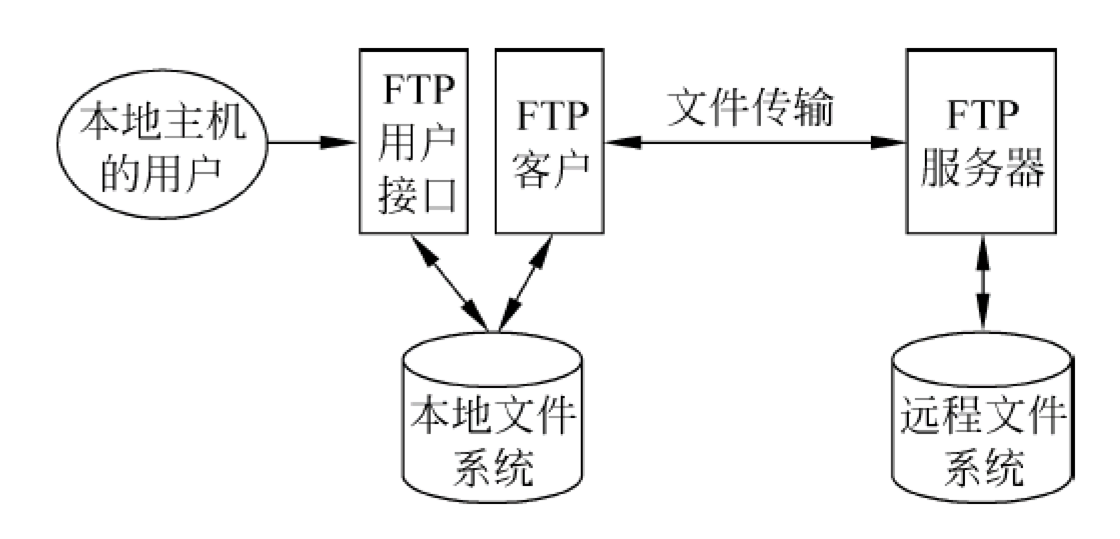


图 8-2 FTP 在本地和远程文件系统之间传输文件

(3) 电子邮件服务。电子邮件(Electronic Mail)简称为 E-mail,是一种通过计算机网络与其他用户进行通信的现代化手段。要想使用 Internet 提供的电子邮件服务,用户必须向提供电子邮件服务的机构提出申请,该机构在其与 Internet 联网的计算机上为用户建立一个电子邮箱,分配一个 E-mail 地址。当用户希望通过 Internet 给某人发送信件时,首先要与提供电子邮件服务的计算机联机,然后将要发送的信件和收信人的 E-mail 地址输入自己的电子邮箱,电子邮件系统自动地将邮件在网上逐站传输到目的地。

(4) 名址服务。名址服务器可以向用户提供对用户名址、计算机名址和 E-mail 地址的查询服务。

(5) 文档查询服务。用户使用 FTP 服务的目的一般是希望在网上找到所需要的文件,并将其下载到自己的计算机上,但用户往往不知道他所需要的文件在哪台 FTP 服务器上。Internet 中有一种被称为文档查询服务器(Archie Server)的计算机,用户只要向这种服务器提供希望查找文件的文件名或文件描述说明中包含的字符串,文档查询服务器就能查找

到存放着所需文件的 FTP 服务器。

(6) 网络新闻服务。网络新闻(Network News)是由一些 Internet 用户为了讨论共同感兴趣的问题而组成的一种逻辑上的用户交流网络。用户新闻组是完全自由参加的组织,参加时无须事先申请,不感兴趣时也无须声明退出。用户的计算机上只要安装了被称为“新闻阅读器”的程序,就可以通过 Internet 随时阅读服务器提供的消息,并能把自己的观点提供给服务器,作为消息在组内发布。

(7) Gopher 服务。Gopher 是基于菜单驱动的 Internet 信息查询工具,它能将用户的请求自动转换为 FTP 或 Telnet 命令。用户通过菜单查找感兴趣的信息资源,无须直接使用 FTP 或 Telnet 命令。Gopher 为 Internet 开创了一种崭新的查询方式,颇受用户欢迎,被看做当今 Internet 类检查工具的先祖。

(8) WAIS 服务。WAIS 服务(Wide Area Information Service,广域信息服务)与前面讲的文档查询服务(Archie Service)不同。文档查询服务是基于文件名的自动搜索服务,而 WAIS 是基于文件内容(关键字)的自动搜索服务。Web 用户可以下载一个 WAIS 客户机程序和一个网关到 Web 浏览器,或者通过远程登录连接到一个公共的 WAIS 客户机来使用 WAIS。由于现在已有丰富的服务器文件和搜索引擎,大多数的 Web 用户会觉得 WAIS 是多余的。然而,图书管理员、医学研究员等一些专业人士需要通过 WAIS 来获得目前 Web 上没有的专业信息。

(9) WWW 服务。下文将对该服务进行详细介绍。

8.1.2　World Wide Web 简介

1. Web 的产生和发展

World Wide Web 简写为 WWW 或 Web,是 Tim Berners-Lee 在 CERN(欧洲粒子物理实验室)发明的。1990 年 3 月,CERN 的 Tim Berners-Lee 建议开发一个超文本系统,以帮助分散在各地的高能物理研究人员能够更有效地共享最新的研究信息,这是一个寻求建立“分散的多媒体系统”的计划。

Web 技术正迅速发展,它不仅提供一种在组织内部进行业务活动的手段,还可促使商家和消费者之间进行交易,并且 Web 能够帮助发展信息资源,这种资源势力强大,商家、学校和 Internet 用户正越来越依赖它。Web 实际上是世界范围内相互联系的文件的大集合。

2. Web 的工作原理

1) 客户机/服务器模式

在 Internet 上,Web 是以客户机/服务器的方式(C/S 或 Client/Server)工作的。所有的 C/S 系统都可以分为 3 个部分:客户机、中间件和服务器。客户机涉及软件的用户前端部分,它常常使用图形用户界面(GUI)。中间件位于客户机与服务器之间,通常对用户是透明的。中间件一般有 3 层:硬件、协议和 API。服务器通常是作为核心的程序或机器,提供对客户机的服务。目前在 Internet 上的服务种类就是前面提到的 9 种服务。当用户需要服务时与客户机的用户界面交互,客户机将这些交互事件序列翻译成服务器能接受的命令,通过中间件发送给服务器。服务器检查并执行这些命令,将结果通过中间件发送回客户机。客户机再将这些结果转换为用户界面的格式提供给用户。

Web 充分发挥了 Internet 和 C/S 两方面的优势,并将两者有机结合,在 Internet 上建立了星罗棋布的各类服务器,使用统一的协议,为遍及世界各地的各类用户提供服务。

2) 超文本模式

Web 在 Internet 上的工作是基于超文本标记语言(HTML)、超文本传输协议(HTTP)和统一资源定位器(URL)的。

超文本(Hyper Text)将以往基于线性方式的信息改为非线性的、非顺序的、联想式的文本组织方式。这种方式更符合人们的阅读习惯。同时用户可以动态地重构链接,以组成新的满足自己需要的阅读文本。

超文本将零星信息组织在一系列离散的节点(Node)之中,并通过链(Link)建立起节点与节点之间的联系,形成由节点-链构成的网状信息结构。超文本的内容不限于文本,可能是图像、图形、声音、动画等多种介质的信息。

超文本可以简单定义为收集、存储和浏览离散信息以及建立和表示信息之间关系的技术。超文本技术不同于传统的计算机技术,它不仅注重信息本身,更注重表示信息之间的关系,用超文本技术建立和浏览信息,主要是根据关系,而不是传统的索引。它可以把任何相关的事物联系到一起。由于信息是按关系组织的,所以整个系统是网状的,而不是树状的,即用超文本技术的信息系统无所谓头尾。建立和浏览信息系统时用户可以任意选择起始点和顺序。

超文本赋予了 Web 一种强有力的功能,它使得用户以同一种方式来访问所有的 Internet 资源。

3) HTML 语言

Web 信息服务系统使用的超文本是用 HTML 语言(Hyper Text Markup Language,超文本标记语言)编写的。当用户使用 Web 服务客户浏览程序通过 Internet 阅读这些超文本时,客户浏览程序负责解释文本中嵌入的 HTML,并按照 HTML 命令将文本中的信息显示给用户。

HTML 语言是使用标准的通用标记语言 SGML(Standard Generalized Markup Language)定义的一种基于在普通正文文件中加入标记(Tag)的方法生成超文本的语言,它允许 Web 文本的作者把信息内容和文本的表现形式分开。由于使用 HTML 语言,Web 文本含有特别的标记说明文本的标记和章节的标题和链接,这就为通过计算机网络搜索 Web 文本提供了方便条件。HTML 语言编写的信息按多级标题结构进行组织,其结构如下:

```
<HTML>
<HEAD><TITLE>标题名</TITLE></HEAD>
<BODY>
        <H1>一级标题名</H1>
         ……Web 页主体
</BODY>
</HTML>
```

4) 超文本传输通信协议 HTTP

Web 在 Internet 上是基于超文本传输通信协议 HTTP (Hyper Text Transfer Protocol)进行工作的。Web 客户机与服务器通过 HTTP 建立连接和完成超文本在

Internet 上的正确传输。HTTP 协议是一种很简单的通信协议，其实现基础是通过网络查询的文件包含着可以实现进一步查询的链接。HTTP 定义浏览器和服务器如何通信并进行信息传输。

HTTP 采用请求/响应过程。客户端浏览器发送一个指定信息页的请求到服务器，服务器收到请求后找出请求页(文件)，并把它送到浏览器，浏览器和服务器必须保持处理请求时的连接。HTTP 通常用来为浏览器传输 HTML 文件，但它也可以传输其他任何文件。如果浏览器不能显示这一文件，它会启动另一个应用程序(帮助程序) 来显示。如果没有帮助程序，它会提示用户将其保存到磁盘上。表 8-1 概括了 HTTP 协议的主要特点。

表 8-1　HTTP 协议的主要特点

通　用　性	支持客户机/服务器模式
简单快速	客户向服务器请求服务时，只需传输请求方法和路径
灵活	HTTP 允许传输任意类型的数据对象
无连接	每次连接只处理一个请求，服务器处理完客户的请求并收到应答后，即断开连接
无状态	协议对于事务处理没有记忆能力

5) SQL 语言

SQL 是 Structured Query Language(即结构化查询语言)的简写。SQL 是高级的非过程化编程语言，允许用户在高层数据结构上工作。它不要求用户指定对数据的存放方法，也不需要用户了解具体的数据存放方式，因此具有不同底层结构的不同数据库系统可以使用相同的 SQL 语言作为数据输入与管理的接口。它以记录集合作为操作对象，所有 SQL 语句接受集合作为输入，返回集合作为输出，这种集合特性允许一条 SQL 语句的输出作为另一条 SQL 语句的输入。所以 SQL 语句可以嵌套，这使它具有极大的灵活性和强大的功能，在多数情况下，在其他语言中需要一大段程序实现的功能只需要一个 SQL 语句就可以达到目的，这也意味着用 SQL 语言可以写出非常复杂的语句。图 8-3 所示为 SQL 语句的基本结构。

SQL 最早是 IBM 的圣约瑟研究实验室为其关系数据库管理系统 SYSTEM R 开发的一种查询语言，它的前身是 SQUARE 语言。SQL 语言结构简洁，功能强大，简单易学，所以自从由 IBM 公司 1981 年推出以来，SQL 语言得到了广泛应用。如今无论是像 Oracle、Sybase、Informix、SQL Server 等这些大型的数据库管理系统，还是像 Visual FoxPro、PowerBuilder 等这些 PC 上常用的数据库开发系统，都支持 SQL 语言作为查询语言。

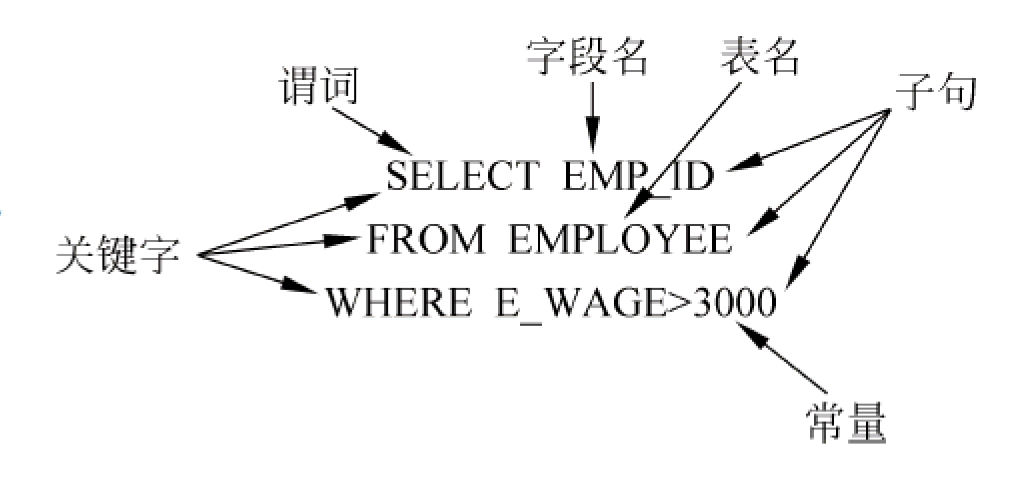


图 8-3　SQL 语句的基本结构

美国国家标准局(ANSI)与国际标准化组织(ISO)共同制定了 SQL 标准。ANSI 是一个美国工业和商业集团组织，负责开发美国的商务和通信标准，同时也是 ISO 和 International Electro technical Commission(IEC)的成员之一。ANSI 发布与国际标准组织相应的美国标准。1992 年，ISO 和 IEC 发布了 SQL 国际标准，称为 SQL-92。ANSI 随之发布的相应标准是 ANSI SQL-92(有时被称为 ANSI SQL)。尽管不同的关系数据库使用的

SQL 版本有一些差异,但大多数都遵循 ANSI SQL 标准。SQL Server 使用 ANSI SQL-92 的扩展集,称为 T-SQL,遵循 ANSI 制定的 SQL-92 标准。

SQL 语言包含如下 4 个部分:

(1) 数据定义语言(DDL)。例如 CREATE、DROP、ALTER 等语句。

(2) 数据操作语言(DML)。例如 INSERT(插入)、UPDATE(修改)、DELETE(删除)语句。

(3) 数据查询语言(DQL)。例如 SELECT 语句。

(4) 数据控制语言(DCL)。例如 GRANT、REVOKE、COMMIT、ROLLBACK 等语句。

SQL 语言包括 3 种主要程序设计语言类别的语句:数据定义语言(DDL)、数据操作语言(DML)以及数据控制语言(DCL)。

6) 统一资源定位器 URL

URL(Universal Resource Locator,统一资源定位器)是一种统一格式的 Internet 信息资源地址表达方法,它将 Internet 提供的各类服务统一编址,以便用户通过 Web 客户程序进行查询。在格式上,URL 分为 3 个基本部分:信息服务类型:、//信息资源地址、/文件路径。

7) Web 浏览器

Web 浏览器是一种功能强大的用于在 Web 上阅读文件的工具,它可以从不同的来源获取文件。Web 站点和信息提供者使这些文件可以被浏览器通过多媒体服务器阅读。此外,Web 浏览器可以使用 FTP,NNTP (Network News Transport Protocol,网络新闻传输协议)等其他方法访问文件。Web 浏览器的常见功能如表 8-2 所示。

表 8-2 Web 浏览器常见功能表

支持标准	HTTP(超文本传输协议)和 HTTPS HTML(超文本标记语言),XHTML(可扩展的超文本标记语言) XML(可扩展标记语言),图形档案格式如 GIF、PNG、JPEG、SVG CSS(层叠样式表),JavaScript(动态网页 DHTML) Cookie,电子证书,Macromedia Flash Java applet,Favicons
基本功能	书签管理,下载管理,网页内容缓存 透过第三方插件(Plugins)支援多媒体
附加功能	网址和表单资料自动完成,分页浏览 禁止弹出式广告,广告过滤

8.1.3 Web 的特点

(1) Web 是图形化的和易于导航的(navigate)。Web 非常流行的一个很重要的原因就在于它可以在一页网页上同时显示色彩丰富的图形和文本。而在 Web 之前,Internet 上的信息只有文本形式。Web 具有将图形、音频、视频信息集合于一体的特性。同时,Web 是非常易于导航的,只需要从一个链接跳到另一个链接就可以在各页各站点之间进行浏览了。

(2) Web 与平台无关。浏览 WWW 对用户的系统平台没有什么限制。无论从

Windows 平台、UNIX 平台、Macintosh 平台还是其他平台，用户都可以访问 WWW。对 WWW 的访问是通过一种叫做浏览器（browser）的软件实现的，例如 Netscape 的 Navigator、NCSA 的 Mosaic、Microsoft 的 Explorer 等。

（3）Web 是分布式的。大量的图形、音频和视频信息会占用相当大的磁盘空间，用户甚至无法预知信息的多少。对于 Web 而言，没有必要把所有信息都放在一起，信息可以放在不同的站点上，只需要在浏览器中指明这个站点就可以了。Web 能使物理上并不一定在一个站点的信息在逻辑上一体化，而在用户看来这些信息是一体的。

（4）Web 是动态的。由于各 Web 站点的信息包含站点本身的信息，信息的提供者可以经常对站上的信息进行更新，例如某个协议的发展状况、公司的广告等，各信息站点通常都尽量保证信息的时间性。所以 Web 站点上的信息是动态的，经常更新的。这一点是由信息的提供者保证的。

（5）Web 是交互的。Web 的交互性首先表现在它的超链接上，用户的浏览顺序和所到站点完全由他自己决定。另外通过 FORM 的形式可以从服务器方获得动态的信息。用户通过填写 FORM 向服务器提交请求，服务器根据用户的请求返回相应信息。

8.2　Web 面临的安全威胁

在网络中，信息系统的硬件、软件和相关数据都需要加以保护，所以网络信息安全本质上就是要使这些部分不会遭到破坏、更改和泄露，从而保证信息系统能够连续、可靠、正常地运行。网络自身存在的脆弱性是网络信息安全问题产生的首要因素。工作场地和环境的恶劣程度，自然灾害的发生以及元器件的老化、失效等因素都对网络的物理方面有重大影响。IPv4 通信协议带来的安全隐患，操作系统和数据库系统存在的安全漏洞，网络安全技术的不完善等因素则对网络的信息安全方面存在重大影响。物理方面的脆弱性较为具体实在，便于人们采取措施强化，而信息安全方面的脆弱性比较复杂，人们难以控制。

客观环境和人为因素则是网络信息安全问题产生的另一个原因。具体的表现有全球信息战的进行、高智商罪犯和黑客的威胁、网络安全意识的缺失和监督管理措施不到位等。

在 Web 技术飞速演变的今天，安全风险达到了前所未有的高度。由于 Web 服务器提供了几种不同的方式将请求转发给应用服务器，并将修改过的或新的网页发回给最终用户，这使得非法闯入网络变得更加容易。许多程序员不知道如何编写具有强效安全性的应用程序，其开发出的应用程序存在安全缺陷。这些安全缺陷被恶意用户利用后很可能会出现灾难性的后果。很多黑客通过服务器、存在缺陷的程序和代码对 Web 应用进行攻击，这些攻击能很直接地通过防火墙等安全措施。Web 应用程序攻击包括 Cookie 毒害、强行浏览、跨站脚本执行、参数篡改、对信息的输入控制、缓冲区溢出、后门、DOS(拒绝服务)攻击等。

1. Cookie 毒害

传统的 Web 应用系统为了支持面向用户的网页内容，通常都使用 Cookie(有时也用其复数形式，即 Cookies)。网站常常将一些包括用户 ID、密码、账户等的 Cookie 存储到用户系统上。通过改变这些值，恶意的用户就可以访问不属于他们的账户。攻击者也可以窃取用户的 Cookie 并访问用户的账户而不必输入 ID 和密码或进行其他验证。

2. 强行浏览

黑客通过改变程序流程，能够强行访问一些正常情况下无法获得的信息和程序，例如日志文件、管理工具以及Web应用程序的源代码。如果Web应用程序没有进行正确的配置，恶意的用户就可能直接访问到带有敏感信息的URL。

3. 跨站脚本执行

一般来说，跨站点编写脚本是将代码插入由另一个源发送的网页之中的过程。利用跨站点编写脚本的一种方式是通过HTML格式，将信息贴到公告牌上就是跨站点脚本编写的一个很好范例。恶意的用户会在公告牌上贴上包含有恶意的JavaScript代码的信息。当用户查看这个公告牌时，服务器就会发送HTML与这个恶意的用户代码一起显示，客户端的浏览器会执行该代码。跨站脚本执行漏洞的成因就是因为CGI程序没有对用户提交的变量中的HTML代码进行过滤或转换。它导致的威胁包括获取其他用户Cookie中的敏感数据、屏蔽页面特定信息、伪造页面信息、拒绝服务攻击、突破外网/内网不同安全设置、与其他漏洞结合、修改系统设置、查看系统文件和执行系统命令等。

4. 参数篡改

参数篡改包括操纵URL字符串来检索用户用正常方式得不到的信息。访问Web应用的后端数据库是通过包含在URL中的SQL调用来进行的。恶意的用户可以操纵SQL代码，从而有可能检索一份包含所有用户、密码、信用卡号的清单或者储存在数据库中的任何其他数据。导致该漏洞的主要原因是Web应用程序没有对客户端提交的参数进行严格的检验。

5. 输入信息控制

输入信息控制包括通过控制由CGI脚本处理的HTML格式中的输入信息来运行系统命令。攻击者如果控制了CGI脚本向另一个用户发送信息的形式，就能将服务器的密码文件邮寄给恶意的用户或者删除系统上的所有文件。

6. 缓冲区溢出

缓冲区溢出是指当计算机程序向缓冲区内填充的数据位数超过了缓冲区本身的容量时，溢出的数据覆盖在合法数据上。理想的情况是，程序检查数据长度并不允许输入超过用户缓冲区长度的字符串。但是绝大多数程序都会假设数据长度总是与所分配的存储空间相匹配，这就为缓冲区溢出埋下隐患。操作系统所使用的缓冲区又被称为堆栈，在各个操作进程之间，指令被临时存储在堆栈当中，堆栈也会出现缓冲区溢出。当一个超长的数据进入到缓冲区时，超出部分就会被写入其他缓冲区，其他缓冲区存放的可能是数据、下一条指令的指针，或者是其他程序的输出内容，这些内容都被覆盖或者破坏掉。如果相关数据里包含了恶意代码，那么溢出的恶意代码就会改写应用程序返回的指令，使其指向包含恶意代码的地址并被CPU编译而执行，这就是内存缓冲区溢出攻击。

7. 调试选项及后门

开发人员常常建立一些后门，并依靠调试选项来排除应用程序的故障，这些安全漏洞经常被留在一些放在Internet上的最终应用中。一些常见的后门使用户不用密码就可以登录或者访问允许直接进行应用配置的特殊URL。

8. 不安全的配置

许多操作系统和应用程序的默认配置是非常不安全的，在启用之前应该被重新配置。这些默认配置包括开放了大量不必要的服务、安装某些不安全的第三方软件、使用默认密码、默认例子程序、不正确的文件访问许可设置等。

9. 管理缺失

许多操作系统和第三方应用软件都存在一些已知漏洞，如果管理员不及时对 Web 站点进行维护，安装已经发布了的软件补丁，这些漏洞就很可能被黑客利用。因为黑客只需要利用简单的漏洞扫描器和大量的漏洞披露网站就可以知道该怎样实施攻击了。

8.3　针对 Web 应用程序漏洞的攻击

Web 应用程序的安全漏洞包括由 Web 站点中的编程错误引起的用户详细信息被暴露、恶意用户执行任意的数据库查询、通过远程命令行访问服务器等非安全行为。漏洞扫描器的基本原理如图 8-4 所示。

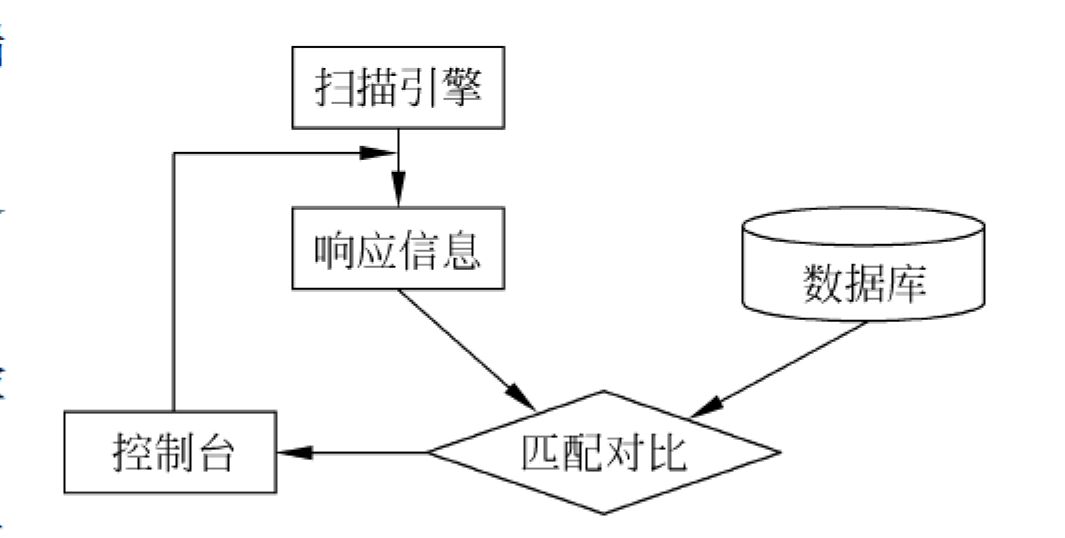


图 8-4　漏洞扫描器的基本原理图

一种常见的基于网络扫描 Web 应用程序漏洞的过程大致分为以下 4 步：

(1) 确定检测目标，主要是确定 Web 服务器的 IP 地址或域名地址。

(2) 从 Web 应用程序漏洞库列表中提取检测漏洞，或者从插件组中选取检测插件。

(3) 发送检测请求，根据响应信息判断是否存在漏洞。

(4) 返回 Web 应用程序存在的漏洞类别。

在确定检测目标后，检测过程如图 8-5 所示。

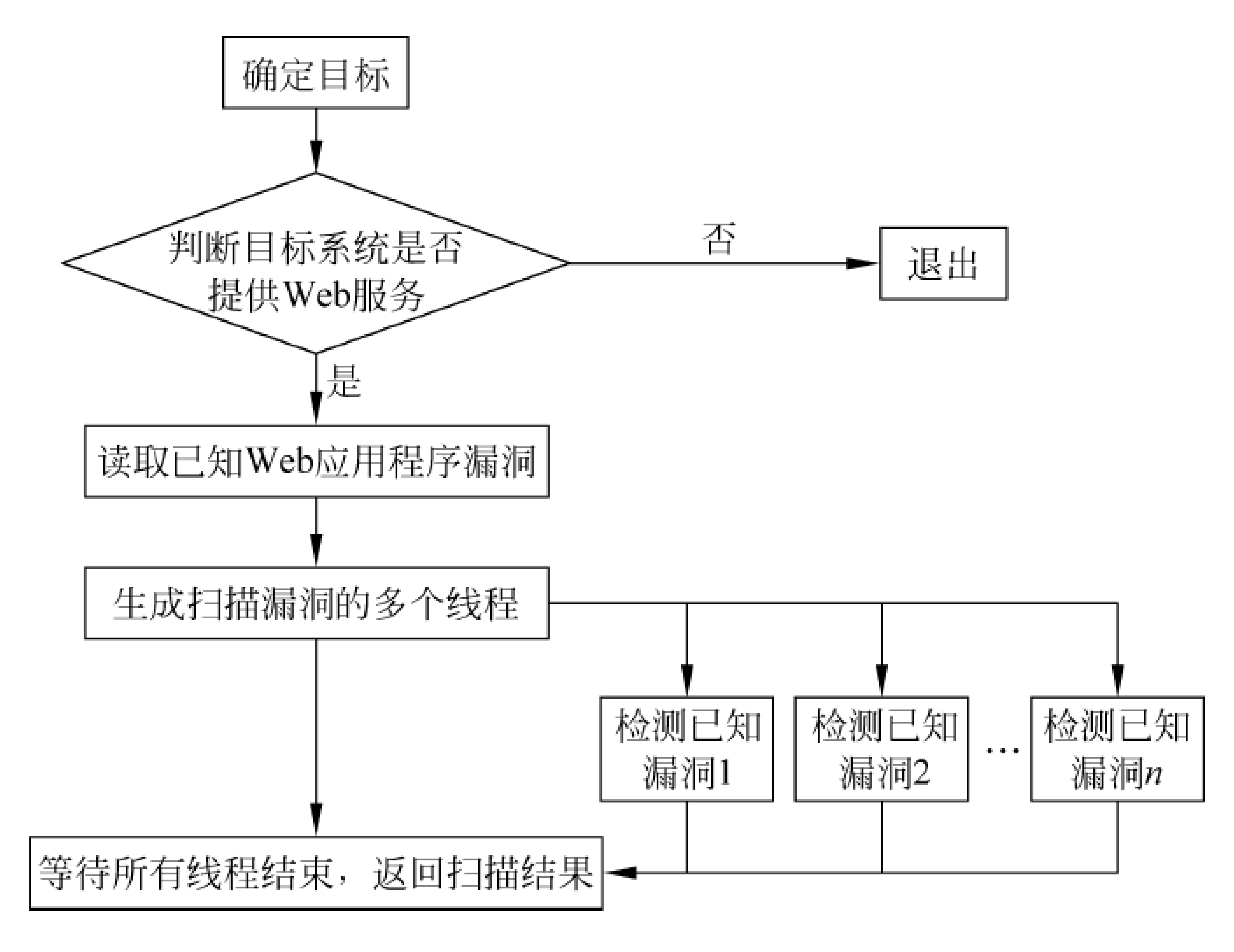


图 8-5　扫描 Web 应用程序漏洞的过程

基于 Web 的应用程序漏洞是错综复杂的，常见的 Web 应用程序漏洞有以下几大类。

1. 物理路径泄露漏洞

提供 Web、FTP 等公共服务的服务器都可能出现物理路径泄露的问题。导致服务器路径泄露有可能是 Web 平台本身、脚本语言解释器、引擎插件、组件、辅助程序等因素。很多情况下 Web 服务器路径泄露漏洞是由 Web 服务器处理用户请求出错导致的，这些用户请求可能是一个超长的请求，或是某个精心构造的特殊请求，或者被请求的文件在 Web 服务器上并不存在。这些请求都有一个共同特点：被请求的文件肯定属于 CGI 脚本，而不是静态 HTML 页面。对于很多 CGI 论坛来说，提交一个不存在的 CGI 文件请求，或者提交一个没有输出的 CGI 文件的请求，请求返回结果将会把网站本身所在的物理路径暴露出来。另一种类型的 Web 服务器路径泄露漏洞是由于 Web 服务器的某些显示环境变量的程序错误输出了 Web 服务器的物理路径。普通用户通过 Web 站点享受 Internet 提供的服务、进行信息交流，黑客则想方设法对 Web 站点进行攻击，对网页进行 SQL 注入、篡改网页、利用 Web 站点传播木马给浏览网站用户、利用 Web 站点获取非法利益等。

2. 源代码泄露漏洞

源代码泄露漏洞的产生是由输入验证错误引起的。以 IIS 为例，IIS 是通过文件扩展名来决定将一个文件内容直接显示出来还是作为脚本进行执行的。对于 ASP 文件，如果请求中的扩展名是.ASP，IIS 可以正确处理。如果将扩展名后面加一个“%52”，IIS 将不认为这是一个 ASP 文件，也就不会执行。但是文件系统会忽略文件名后的“%52”，会正确找到文件。

3. 目录遍历漏洞

目录遍历攻击指的是恶意用户找到受限文件的位置并且浏览或者执行它们。攻击者浏览受限文件，如果读取了密码文件就会破坏隐私甚至引发安全问题。另一个严重的问题是，攻击者执行受限的文件就可以根据自己的意愿来控制或修改 Web 站点。

目录遍历主要使用猜测文件是否存在的方法进行。

假如有 URL：http://www.test.com/article.asp?id=Ql.htm，首先可以推断出 HTML 文件 Ql.htm 在 Web 中是允许用户读取的，应用程序就正在载入文件，并且将其中的内容显示给用户。其次将 URL 地址简单地改为：http://www.test.com/article.asp?id=Q2.htm。如果存在同名的文件 Q2.htm，并且服务器对于文件的访问权限没有进行限制的话，就可以看到其中的内容了。可以通过这种猜测的方法，获取在 Web 服务器上本不该允许用户访问的文件。

4. 执行任意命令

在 Web 发展的早期，程序员利用 C 语言或者 Perl 语言编写的 Web 应用程序存放于 UNIX 服务器上。执行任意命令攻击的思想就是运行自己输入的命令，而不是按照开发人员预期的那样执行某个指定的程序。攻击目标是发送到服务器上的操作系统命令或可执行程序的用户输入。这些输入域的一部分会常出现在应用程序的参数中，而这些参数可能是页面中包含的一些被引用的文件，也可能是其他程序的参数。任何需要操作系统协助才能完成的功能如创建用户、修改文件和收集主机数据等都需通过某种接口和外部的程序进行交互。如果这些程序中的任意一个不能对发送的数据进行验证，就会是系统中很容易受到

攻击的弱点环节。

5. 缓冲区溢出漏洞

缓冲区溢出是针对 Web 应用最严重的一种攻击，当程序无法检查正在处理的数据输入量时就有可能发生缓冲区溢出问题，如果输入的数据量超过了程序为其分配的内存空间的大小，它就会侵占其他程序堆栈的内存空间，这些内存中原有的其他数据就会被覆盖。大多数情况下，这些被覆盖的数据会导致软件崩溃。

6. 拒绝服务攻击

拒绝服务(DOS)攻击的思想是代码执行总是需要时间的，每次由 Web 服务器、应用程序或数据库调用的函数，其执行过程总要耗费一定的处理器周期。如果执行过程需要很长时间，并且操作系统无法将其切换到其他程序上，服务器就会被束缚在为一个请求进行服务的过程中。因此，现在多数的 Web 服务器都采用了多线程，系统中同时存在多个任务，如果其中的某一个任务占用了过长的时间影响了其他任务的执行速度，操作系统就可以进行切换执行另外的任务。这就存在一种潜在的拒绝服务(DOS)攻击，如果将大量的请求快速提交到 Web 服务器上，并且都是很多耗时的服务，就可以阻止其他用户正常访问 Web 站点。

7. CGI 漏洞攻击

CGI 程序是交互性的，它允许用户把自己的数据按照一定的格式发送给服务器，然后由服务器对其进行解释，之后再把解释的结果传给用户，如果用户提交了一些非正常的数据，那么服务器在解释这些数据时可能会绕过 IIS 对文件名所做的安全检查，在某些条件下攻击者可以执行任意系统命令。

8. 跨站脚本攻击

利用页面进行 XSS(Cross Site Scripting，跨站脚本)攻击的方法主要有两种：

(1) 攻击者最常检查的 XSS 漏洞就是那些让用户输入信息以供其他浏览此页的用户进行查看的地方，例如留言、评论和用户登录框。当其他用户浏览此 Web 页面时，应用程序就从存储单元中搜集数据，并且将其显示出来。采用这种回显静态数据的方法通常是有害的，攻击者可以用脚本代替静态的数据。在这种情况下，攻击者通常将脚本直接输入到被攻击的站点的表格域中。此时，脚本代码被留在被攻击的服务器上，当其他用户访问这个页面时，这些脚本就开始执行。

(2) 将脚本嵌入到 URL 地址的 CGI 参数中。这种情况下，攻击者可以通过电子邮件将一个链接发送到潜在的受害者，当攻击者单击此链接时，页面被下载，但其中的内容被嵌入在 URL 中的脚本修改了。这种情况无须将脚本存放在受攻击的站点中，因为脚本是在用户单击了某个被篡改过的链接时的页面载入过程中才完成执行的。HTML 支持很多种将脚本嵌入页面中的机制，最常见的就是使用<script>…</script>标签，其他标签还有<html>、<body>、<embed>、<frame>、<frameset>、<applet>、<iframe>、<img>、<Layer>、<meta>、<object>以及<style>，这些都可能会引发跨站脚本漏洞问题。

跨站脚本攻击过程示意图如图 8-6 所示。

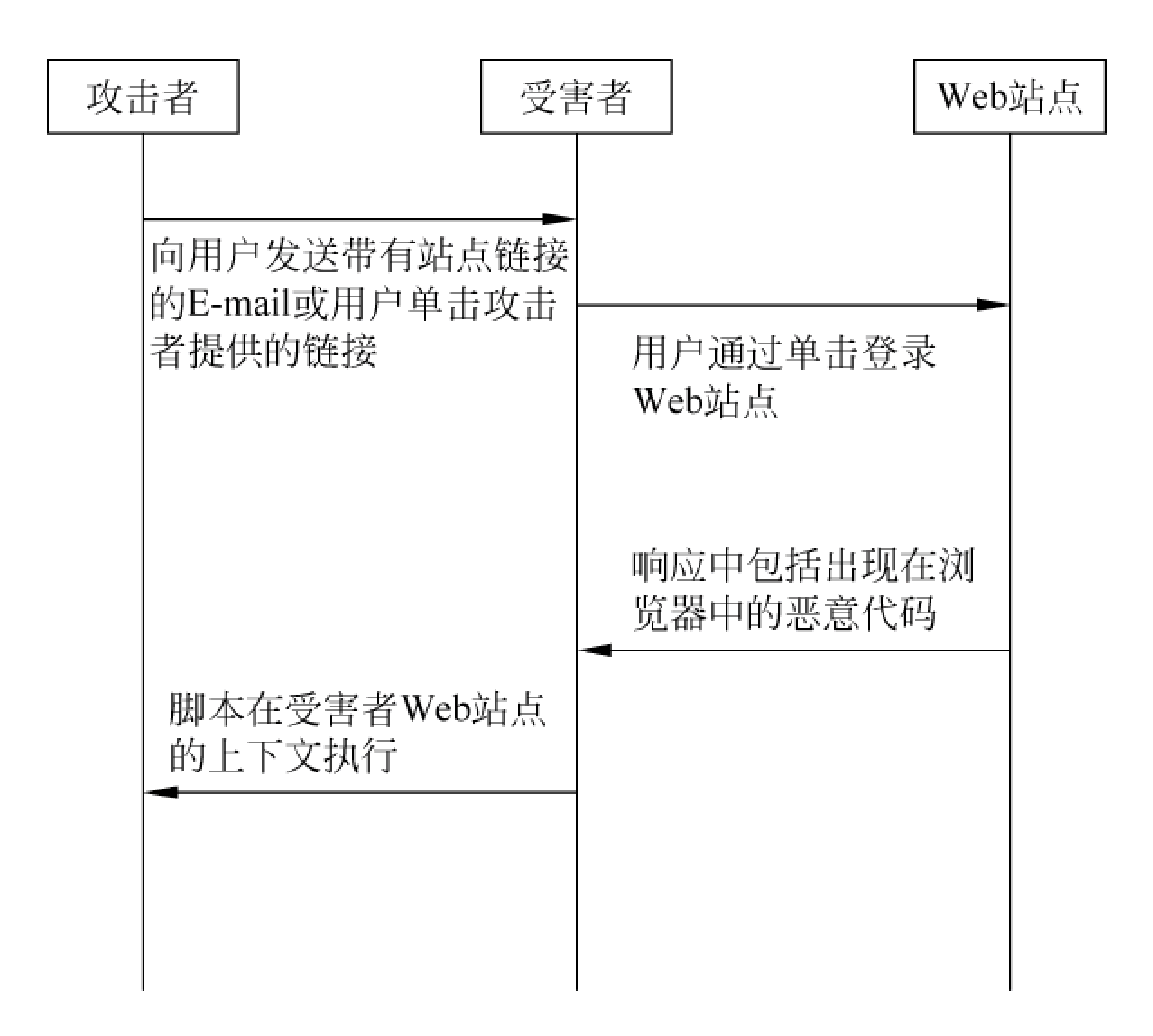


图 8-6 跨站脚本攻击过程示意图

9. SQL 注入攻击

SQL 命令就是前端 Web 和后端数据库之间的接口，使得数据可以传输至 Web 应用程序，也可以从中发送出来。很多 Web 站点都会利用用户输入的参数动态地生成 SQL 查询要求，攻击者通过在 URL、表格域或者其他输入域中输入自己的 SQL 命令，以此改变查询属性骗过应用程序，从而对数据库进行不受限的访问。SQL 注入是其中一种传播范围广、危害严重的功能攻击方式。为了利用 SQL 注入漏洞，攻击者必须找到一个参数用来传递数据，如果 Web 网络应用程序有这类漏洞，就可能把这个参数传输到操作数据库的 SQL 语句中，应用程序使用该语句操作数据库就可能导致信息泄露、数据丢失、记录篡改等危害。SQL 注入攻击示意图如图 8-7 所示。

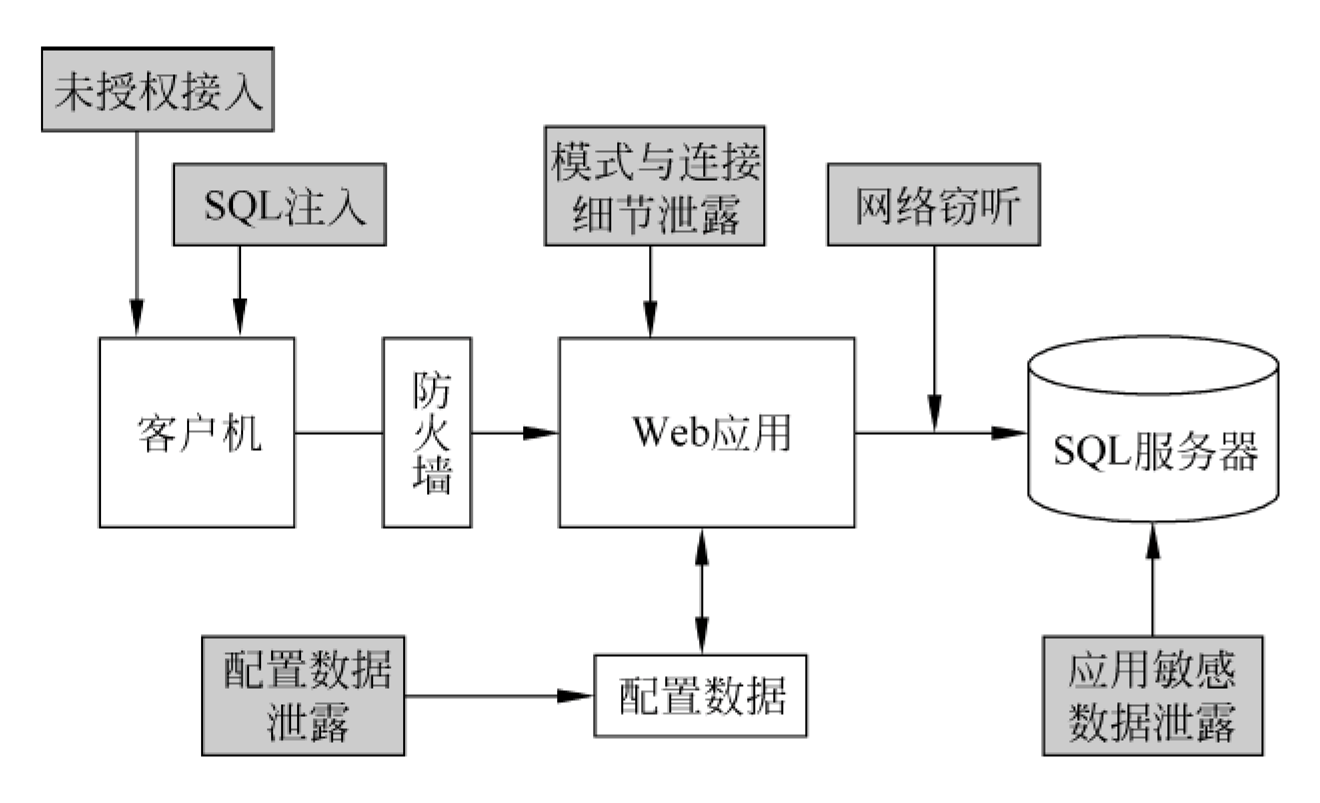


图 8-7 SQL 注入攻击示意图

10. 未验证的输入

Web 应用程序一般是根据 HTTP 请求中用户的输入决定如何响应，黑客能够利用 HTTP 请求中的任何一部分，包括 URL、请求字符串、Cookie 头部、表单项或隐含参数传递

代码来发动攻击。一些网站使用过滤器过滤掉恶意输入，但是由于输入信息存在着各种各样的编码方式，几乎所有的 HTTP 输入都可以被表示成多种形式，使用编码技术可以绕过 Web 应用程序的验证与过滤机制。很多网络应用程序只使用了客户端机制来验证输入。但是，客户端的验证机制可以被简单地绕过去，这就使得网络应用程序直接面对那些恶意的输入参数而毫无防备。攻击者可以使用简单的工具例如 telnet 来生成他们的 HTTP 请求，因此必须要设置服务器端的验证来防备那些恶意参数的攻击。

11. 被破坏的认证和会话管理

授权和会话管理包括处理用户授权和管理激活态的会话的所有方面。授权是这个过程的关键方面，但是即使是固有的授权认证也有可能遭到含有漏洞的信任管理功能破坏。这些固有的授权认证包含修改密码、忘记密码、记住密码、账户更新和其他更新功能。在网络中，通常的用户授权包括 UserID 和密码的使用。大多数的账户和会话管理漏洞可能破坏用户或系统管理员账户。网络应用程序必须建立会话来跟踪每个用户的请求数据流，HTTP 不提供这方面的能力，所以网络应用程序必须自己创建。通常而言，网络应用程序环境提供了一个会话，但是大多数的开发人员更希望建立它们自己的会话标记。无论在哪一种情况下，如果会话标记没有能够妥善保存，攻击者就可能截获一个处在激活态的会话并且假扮成这个用户的身份标识。

12. 不当异常处理

不当的异常处理可能给网站带来各种各样的安全问题。最常见的问题就是向用户显示内部出错信息，如果这些出错信息不加选择地展现到用户面前，就可能公开了本不应该公开的细节。这些细节可能为黑客提供网站潜在漏洞的重要线索，这些信息也会干扰到正常用户。在正常操作时，Web 应用程序也会生成一些出错情况，例如内存不够、空指针异常、系统调用失败、数据库不存在或网络超时等。这些错误必须被一个既定的严密方案正确处理，从而为用户提供一份有意义的出错信息，为 Web 应用维护人员提供诊断信息，而不是为攻击者提供有用的信息。出错信息不能提供过多细节，因为出错信息的差异也会把该 Web 应用是如何工作的等重要信息暴露出来，并且暴露了那些出错信息背后的隐含意义。例如，当一个用户试图访问一份他无权访问的文件时，通常的出错信息会给出“访问被拒绝”的提示，这种提示很可能会暴露这个文件是否存在或这个 Web 应用的目录结构等信息。

13. 不安全的存储

大多数 Web 应用程序都需要存储敏感的信息，这些信息存储在数据库或文件系统的某个位置上。人们通常使用加密技术来保护这些敏感信息，然而虽然加密技术的使用很容易，开发人员还是常常在加密技术和应用程序的结合上出错。经常出错的几个地方包括：

(1) 未对关键数据进行加密。

(2) 密钥、证书和密码的不安全存放。

(3) 在内存中不恰当地保存关键信息。

(4) 不当的随机资源。

(5) 不当的算法选择。

(6) 一种新开发的加密算法。

(7) 当密钥更改或者其他必备的维护过程发生时，无法提供最新的技术支持。

这些薄弱环节会给系统带来严重的安全隐患，受保护的资源也可能因为这些薄弱环节遭受到严重破坏。

14. 不安全的配置管理

Web应用程序服务器的配置对于Web应用程序的安全起到了关键作用。许多应用程序服务器提供了Web应用程序能够使用的服务，例如数据存储、目录服务、邮件和信息处理等。没有合理配置的服务器很可能导致各种安全问题。许多服务器的配置问题影响了安全性，它们包括：

(1) 服务器软件漏洞或者错误的配置允许列出目录和进行目录遍历。

(2) 没必要的默认、备份或者例子文件。

(3) 服务器软件未打补丁的漏洞。

(4) 不当的文件和目录访问权限。

(5) 没有必要的服务，包括内容管理和远程管理。

(6) 使用默认密码和账户。

(7) 被激活的、可以被访问的管理或者调试功能。

(8) 使用默认证书。

(9) 通过外部系统的不正确授权。

(10) 错误配置的SSL证书和加密设置。

8.4 Web应用程序的安全漏洞检测

Web应用程序是指在应用程序服务器的基础上开发的各种应用程序。有时候，虽然攻击者无法击溃Web服务器，却可以利用Web应用程序的漏洞来攻击一个系统。

8.4.1 认证机制漏洞检测

根据Web应用程序要求的不同，可以使用基于HTTP的认证、基于表单的认证以及Microsoft Passport。无论是哪种认证方式，应用程序都会要求用户输入用户名和密码。当然也存在其他形式的基于Web的认证以提供更强的安全性。针对这些认证机制，可以用自动化的攻击方式如猜测密码、破坏Cookie或旁路认证等来绕开认证机制。

1. 密码猜测攻击

一个认证系统，如果在认证协议的选择上或者它的实现上没有任何缺陷，那么它最脆弱的地方就是用户密码的选择。由于许多应用程序服务器总是使用一些熟知的用户名和密码，这就使得用自动的方式来实现密码猜测成为可能。表8-3列出了用于密码猜测攻击常见的用户名和密码。

2. 窃取Cookie

Cookie通常都包含与认证有关的敏感数据。如果Cookie包含了密码或会话ID，那么窃取该Cookie就是对Web站点成功的攻击。用来窃取Cookie的最常用的技术是脚本注入和窃听。

表 8-3 密码猜测字典

用户名猜测	密码猜测
[NULL]	[NULL]
root,administrator,admin	[NULL],root,administrator,admin,password,[公司或组织名]
operator,Webmaster,backup	[NULL],operator,Webmaster,backup
guest,demo,test,trial	[NULL],guest,demo,test,trial
member,private	[NULL],member,private
[公司或组织名]	[NULL],[公司或组织名],password
[已知的用户名]	[NULL],[已知的用户名]

3. 结合 SQL 注入绕过登录表单

很多 Web 站点都使用数据库来保存密码,并使用 SQL 来查询数据库以验证认证证书。如果应用程序没有执行有效的输入验证,则可以采用在用户名之后加上 SQL 注释声明来逃过对密码的检查。

密码猜测攻击可以针对几乎所有类型的 Web 认证来执行。

8.4.2 授权机制漏洞检测

程序开发人员在编写 Web 应用程序的时候经常犯的一个错误就是使用了不正当的授权。普通用户在登录系统之后,应该被禁止访问其他用户的信息,而不正当的授权会使普通用户通过各种手段来提升自身的权限,例如查看其他用户的信息,甚至得到更高级的管理权限。修改查询字符串和修改 URL 是两种常见的攻击授权的注入方式。

1. 修改查询字符串

查询字符串是 URL 中问号后面一些用来传递变量的额外的比特位,用于在客户和服务器之间传递变量,一般是一串由"&"符号分割的名称列表。

例如,http://www.test.com/index.aspx? id=11&admin=false。对此类 URL,攻击者可以尝试改变每个参数的值(例如将 false 改为 true)来实现水平权限提升和垂直权限提升。如果应用程序的授权机制不完善,这种攻击就有可能成功。

2. 修改 URL

这个被修改的 URL 指向一个本来不允许用户访问的资源:

例如,http://www.test.com/new12345.aspx。

攻击者可以尝试访问以下的 URL 来考验应用程序的授权机制:

例如,https://www.test.com/new12346.aspx。

在实际的错误注入过程中,可以对 URL 指向的资源文件的扩展名进行分析,然后尝试用类似的值来访问这个文件。

8.4.3 输入验证漏洞检测

输入验证攻击就是通过提交应用程序没有意料到的数据进行攻击。这些数据可能是对登录窗口的 SQL 注入字符串,可能是一次跨站脚本攻击字符串,也可能是一个能导致应用

程序的缓冲区溢出的超长的字符串。这些漏洞的成因是应用程序中没有正确的验证机制。

1. 应用程序缓冲区溢出漏洞

虽然原理和 Web 服务器的缓冲区溢出攻击相同,但注入超长字符串的位置是 URL 请求中的参数值,这种攻击很容易用工具实现自动化。

2. 用圆点(../..)遍历应用程序的目录

原理和 Web 服务器的目录遍历攻击相同,但目的是为了发现应用程序准确的目录信息和一些关键的文件。把可能包含重要信息的文件名放入一个列表中,然后用圆点和斜杠不断地尝试请求这些文件,以达到测试这类漏洞的目的。

3. 跨站脚本攻击

应用程序的搜索引擎是跨站脚本攻击的首选对象,例如,可以构造一个这样的 URL:

http://www.test.com/research/research.asp?quest=<script>alert('XSS')</script>

若服务器返回一个值为 XSS 的确认对话框响应,就说明 Web 应用程序存在跨站脚本漏洞。

4. 边界检查漏洞

当给 URL 中的某个参数(例如 UserID)一个临界值时,有些应用程序往往会返回一些奇怪的错误,这些错误可能会泄露应用程序的重要信息。例如,若应用程序仅仅接受一个 8 位的 UserID 的值,那么它可接受的值的范围就是 0～255,当用户发出一个 UserID 为 256 的请求时,它会把这个请求当成 UserID 为 0 的情况来处理,这种漏洞也叫回绕错误。对这样的参数类型,可以尝试的值还有 0、-1、255、257、65536、65535 等。

5. 搜索字段的百分号漏洞

百分号(%)通常作为 SQL 语句或者搜索引擎中的通配符。在搜索字段中输入百分号可能会得到整个数据库的内容,或者产生一些错误信息,这些信息可能会让攻击者分析出应用程序的有用信息。

8.5 IIS 和 ASP 技术构造 Web 站点

IIS(Internet Information Server)是 Microsoft 公司提供给 Internet 或 Intranet 的文件和应用服务器。ASP(Microsoft Active Server Pages)则是 Microsoft 公司推出的 Web 应用程序开发技术,提供将脚本语言(VBScript、JavaScript 等)集成到 HTML 网页并作用于服务器端的一种脚本编写环境。ASP 技术是目前比较流行的 Web 服务器和数据库服务器之间的中间件技术。Microsoft 公司将 IIS 和 ASP 结合在一起,用于构建功能强大的 Web 站点。

8.5.1 IIS 自身的安全防护

作为 Web 服务器,IIS 检验登录用户是否合法的过程包含 5 个步骤:IP 地址检查、用户许可检查、IIS 许可检查、自定义身份验证和 NTFS 文件系统许可检查。只有当 5 个步骤的

检查全部通过时,该登录用户才会被允许访问所请求的资源。IIS 的安全设置正是融合在这 5 个步骤之中。

1. IP 地址限制

IIS 能够授予或拒绝特定 IP 地址对其访问,在 IIS 4.0 中,这种授予或拒绝可以细化为对 Web 站点、虚拟目录、目录和文件的访问控制。在 Web 应用系统中,可以根据每个文件或目录的重要性分别设置对 IP 地址的限制。

2. 用户许可检查

在 Windows NT 操作系统的用户账户安全性背景下,每个来自于浏览器的 HTTP 请求都运行于 IIS 之上,IIS 在一个执行线程中执行请求。在该 HTTP 请求执行期间执行的操作都受限于在 Windows NT Server 中授予的用户账户的权限。

IIS 支持 4 种 Web 身份验证模型:匿名(Anonymous)、基本(Basic)、Windows NT 请求/应答(Challenge /Response)和客户凭证映射(Client Certificate Mapping)。Windows NT 只接受合法的用户,不接受匿名用户。但是由于 Internet 是一种极端的匿名,其中为访问者提示用户名和密码的 Web 站点非常少,因此 IIS 创建了 IUSR-ComputerName 账户。这个账户是在安装 IIS 时自动产生的,在本地计算机上生成一个随机密码,专为 Internet 匿名用户使用,同时被授予本地登录用户的权力(匿名用户访问也可以被重置为其他任何合法的 Windows NT 账户)。Windows NT 请求/应答身份验证可以用于限制访问 Web 服务器上的某些重要部分,尤其是在用户拥有 Windows NT 域账户的 Internet 环境中。在 Windows NT 请求/应答身份验证中,浏览器首先使用来自域登录台的当前用户加密身份凭证。如果这些凭证被拒绝,Windows NT 请求/应答身份验证将通过对话框提示用户输入用户名和密码。因此,利用 IIS 的用户许可检查将某些资源的访问权限限定为必须是特定的 Windows NT 的合法用户,禁止匿名访问,以此来保护 Web 应用程序的重要部分。

3. IIS 许可检查

一旦用户被授予访问权,服务器就检查 URL 和请求类型,并检查许可和 SSL 客户身份验证凭证。所谓的许可是指对于 WWW 服务,请求能够指示读、写、执行或脚本行为。一个合适的 WWW 虚拟目录必须有适当的许可授权,否则 WWW 服务将返回一个错误提示"403.X: Access Forbidden",其中的 X 指代访问尝试的类型。

虚拟目录的许可控制有读、写、日志访问、目录浏览、索引。默认设置是读、日志访问、索引该目录。如果一个虚拟目录在一个 NTFS 驱动器上,目录的设置就必须与虚拟目录的设置相匹配。如果不匹配,限制最严格的设置集将发挥作用。例如,如果给一个目录写许可,而同时只给某一个用户组 NTFS 下的读访问许可,那么这些用户就不能写文件到该目录下,因为写许可更加严格。

IIS 还支持对文件操作的许可控制。这些许可是:

None(禁止执行):不允许任何程序或脚本在这些目录下运行。

Script(脚本运行):允许映射到脚本引擎的应用在没有执行(Execute)许可设置的情况下在此目录下运用。

Execute(允许执行):允许任何应用在该目录下运行。

ASP 文件应和 HTML 文件保存在不同的目录下。存放 ASP 文件的虚拟目录要设置

为可执行，但不可被读取。保存 HTML 文件的虚拟目录则必须设置为可读取，否则这些 Web 页将无法显示。

4. 自定义的身份验证

自定义的身份验证意味着必须创建自己的身份验证机制。在系统中最常用的方法是执行一个用户名称和密码的数据库查询。

5. NTFS 文件系统许可

通过使用验证身份方式的用户安全性背景，IIS 可以获得对特定资源(基于 URL)的访问权。每一个合法的 Windows NT 用户账户都有相应的对资源的访问权限。匿名访问是典型的 IUSR-ComputerName 账户。如果前面的身份验证已经被执行，它将是合法的 Windows NT 用户账户，同样也有相应的权限。IIS 的安全登录检查在很大程度上依赖于 NTFS 文件系统，所以最好将虚拟目录都建立在 NTFS 驱动器上。

8.5.2 ASP 的安全编程

ASP 编程中可以使用必要的安全措施以提高站点的安全性。这些必要的安全措施主要有对密码的保护、对 Session(会话)对象的有效利用和控制页面的缓存等。

1. 密码的保护

密码是系统中极敏感、关键的信息。密码的提交最好不要采用明文形式，以防被截取。简单的做法是对密码、用户名进行加密后再传输，服务器接收后再进行解密。也可以利用安全套接层(SSL)协议，通过公共密钥和对称性加密提供非公开通信、身份验证和消息集成。这样，客户和服务器可以在一种防窃听、防干扰、防消息伪造的方式下通信。

2. 对 Session 对象的利用

使用 ASP 内置对象之一的 Session 对象，编程者可以存储任何的数据，包括已实例化的对象。Session 对象使用一个字符串(或关键词)作为索引来存储和检索数值。ASP 使用 HTTP Cookies 识别带有唯一 Session 的用户会话。一旦 ASP 会话开始，ASP 就响应带有设置 Cookie HTTP 头的用户请求，以后每个浏览器的请求都可用会话 ID Cookie 来识别 ASP 会话的状态。Session 对象中存储的变量在用户应用页面间跳转时不会消失，在访问页面的应用中将一直被保留。每个用户在一次会话中都有一个唯一的 Session ID，ASP 使用该值为客户机连接检索正确的 Session 对象。

在应用系统中，编程者通常要控制用户只能按照导航来访问各个 Web 页，不允许直接利用 URL 链接访问资源，更不允许用户直接跳过密码和用户名的检查页面。一个常用的控制方法就是利用 Session 对象。当用户通过安全登录的检查后，就把 Session 对象的 Session ID 属性作为一个 Session 变量存储起来。每当用户试图导航到要求有效链接的页面时，就可以比较当前的 Session ID 与存储在 Session 对象中的 ID。如果它们不匹配，或如果 Session 变量是一个参数值，就可以采取适当的行动拒绝访问。同时，要注意提供用户安全退出的按钮或链接，编程结束这次会话。

3. 控制页面缓存

浏览器一般都有缓存 Web 页面的功能。若用户在 Internet Explorer 浏览器中设置了

“浏览网页时首先查看本地缓冲里的页面”,则当浏览某一网页而本地缓冲中已有该内容时,浏览器会自动浏览缓冲区里的页面,直到该缓冲页面过期后浏览器才会向 Web 服务器去请求新页面。对于合法的用户,这个功能可以加快浏览 Web 页面的速度,而对于不合法的用户则提供了一个越权浏览机会。所以,对于应用系统中重要的 Web 页面,要禁止页面缓存,强制浏览器每次向 Web 服务器请求新页面。当然,这样做无疑会降低应用程序的速度。可采用如下方法:

(1) 在 ASP 文件的首部添加语句:% Response. expires＝0%。Response 对象的 expires 属性表示页面经过多长时间过期,如果设置为 0,则表示立即过期。

(2) 利用 META 标签:<meta HTTP-EQUIV＝“expires”CONTENT＝“0”>。

8.6　防火墙技术应用于 Web 站点的安全

8.6.1　防火墙的功能

最直接的防火墙的利用是创建一个内部站点,它仅能由局域网中的机器访问。这种情况下只需把服务器放在防火墙内部即可,如图 8-8 所示。但从用户组织的整体安全性考虑,最安全的方法是完全将 Web 服务器放到局域网外面,如图 8-9 所示。这种配置方法虽然会使 Web 服务器遭受被攻破的危险,但即使它被攻破也不会破坏内部网络的安全性。

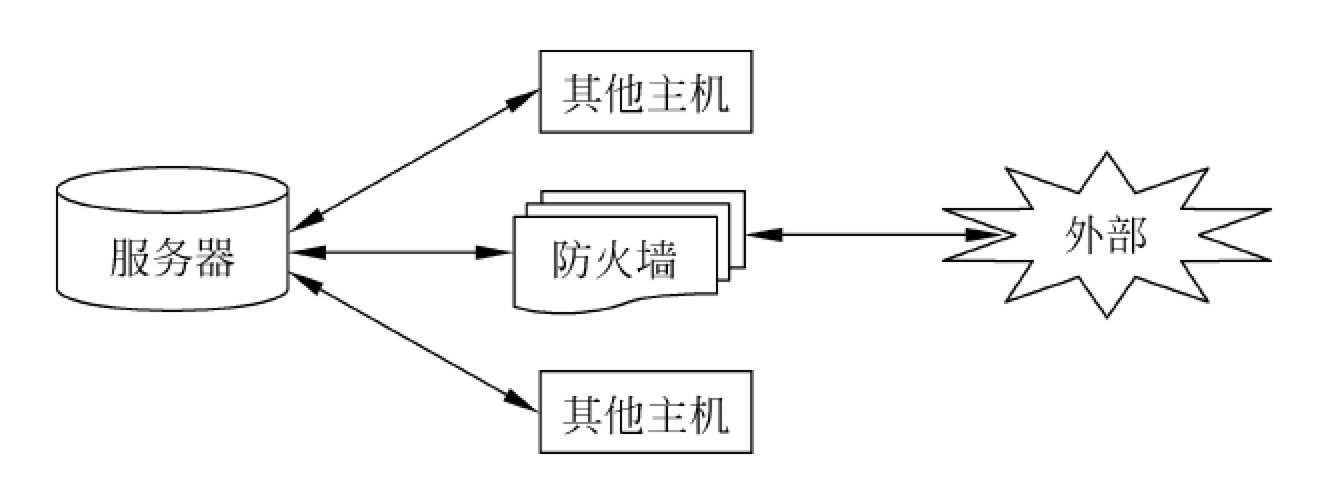


图 8-8　Web 服务器在防火墙内部

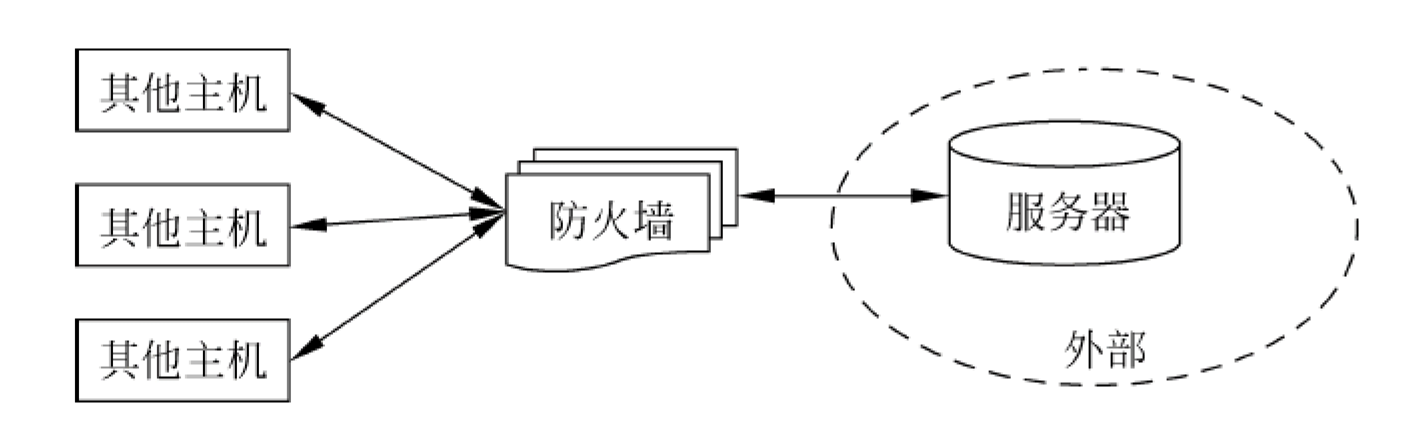


图 8-9　Web 服务器在防火墙外部

8.6.2　代理服务器

代理服务器(有时被称为应用级网关)是一个应用程序,用于协调被保护的网络与 Internet 之间的数据交换,有的代理服务器本身就是一个防火墙软件。代理服务器通常用于替代基于路由器的数据交换控制器,防止数据直接在网络间交换。许多代理程序包含额外的用户认证的日志和支持。由于代理程序必须能够理解使用的应用协议,它必须实现针对协议的安全性,必须专门针对该协议开发出相应的代理软件。

现在流行的代理服务器是 TIS Internet Firewall Toolkit(FWTK),它支持 Telnet、Rlogin、FTP、X Windows、HTTP/Web 和 NNTP/Usenet 新闻。

SOCKS 是一个通用的代理系统,它可以被编译成一个客户端应用程序来使它通过防火墙。其优点是使用简单,缺点是不支持额外的日志和针对协议的日志。

8.6.3 Internet 和防火墙的关系

防火墙作为被保护网络与 Internet 之间的通道,与 Internet 的关系有如下几点:

1. 使 Web/HTTP 通过防火墙

使 Web/HTTP 通过防火墙通常有以下 3 种方法:

(1) 如果使用屏蔽路由器,可以运行通过路由器建立向外的连接。

(2) 使用支持 SOCKS 的 Web 客户程序,并在防火墙上运行 SOCKS。

(3) 在防火墙上运行某种支持代理的 Web 服务器。TIS Internet Firewall Toolkit(FWTK)软件包含一个称为 http-gw 的代理,它支持 Web、gopher/gopher+和 FTP 的代理。CERN httpd 也支持代理。同时,许多 Web 客户程序本身也支持代理服务器,例如 Netscape、Mosaic、Spry 和 Chameleon 等。

2. 使 FTP 通过防火墙

使 FTP 通过防火墙的方法是使用如防火墙工具箱(firewall toolkit)中 ftp-gw 那样的代理服务器,或者通过允许连接到内部具有一定限制的端口区域。

FTP 客户程序会被修改成将数据端口连接到对应的端口区域内,这需要修改主机上的 FTP 客户程序。有时,若用户不希望使用 FTP 下载,就可声明 FTP 为不可用,并令其他用户通过 Web 来下载文件。此外,还可以使用 FTP PASV 选项来指示远程的 FTP 服务器允许客户进行连接。该方法的前提是远程系统上的 FTP 服务器支持该选项。

3. 使 Telnet 通过防火墙

通常通过使用如防火墙工具箱中 tn-gw 那样的代理服务器,或者简单地配置路由器,则路由器允许通过建立的屏蔽规则方法建立外出的连接。应用代理应当是在堡垒主机上运行的独立代理的形式,或是 SOCKS 服务器和被修改的客户的形式。

4. 使 Finger 通过防火墙

许多防火墙仅允许来自信任主机的 Finger 请求,这就说明 Finger 请求的形式是:fingeruser@host.domain@firewall,这种方法只在标准的 UNIX 版本的 Finger 下才能工作。

控制访问服务并将它们限制到特定的主机上,这部分工作是由防火墙工具箱中提供的 tcp_wrapper 或 netacl 来完成的。由于并不是所有的 Finger 服务器都支持 user@host@host 形式的 Finger 请求,因此该方法并不能在所有的系统上运行。

5. X Windows 通过防火墙的问题

X Windows 是一个很有用的系统,但是它存在着一些严重的安全漏洞。远程用户可能欺骗一个 X Windows 系统获取其所有键盘输入信息。虽然目前有一些方法来提高 X Windows 的安全性,但攻击者仍然可以轻易地攻破系统。

很多防火墙禁止所有的 X 数据交换，有的允许通过一定的应用代理来交换 X 数据。用户可以使用防火墙工具箱的 x-gw 或 Telnet 代理。当存在向虚拟 X 服务器（防火墙上的 X 服务器）发送的 X 连接请求时，用户端就会弹出对话框，询问用户是否要允许这个连接。

思　考　题

（1）什么是 Web 的安全性？

（2）Web 的特点有哪些？

（3）Web 面临哪些安全威胁？请简要说明。

（4）什么叫网站漏洞？怎样去找到网站漏洞？

（5）Web 应用程序的漏洞检测包括哪些？它们是如何进行工作的？请进行说明。

（6）Web 应用程序的安全漏洞检测有哪些？它们对 Web 站点的安全起到些什么意义？

（7）什么叫 XSS？它有哪些危害？

（8）网站被 XSS 攻击了，该怎么办？

（9）对于 XSS 攻击，是否可以通过禁止脚本执行来防御？

（10）什么是 SQL 注入？它有哪些危害？

（11）对于 SQL 注入攻击，是否可以通过禁止 SQL 语句执行来防御？弱点检测和漏洞修补是否可以完全防止？

（12）运用 IIS 和 ASP 技术对 Web 站点的建立有什么较大的突破？

（13）请简要阐述防火墙对 Web 站点安全的重要作用。

参 考 文 献

［1］ Jeff Forristal. Web 应用程序的黑客防范. 北京：机械工业出版社，2002.

［2］ 国家互联网应急中心. 中国互联网网络安全报告（2010 上半年），2010.

［3］ Liang Ying，Wang Hui-Qiang，Lai Ji-Bao. Quantification of Network Security Situational Awareness Based on Evolutionary Neural Network Maehine Leaming and Cybemeties. International Conferenee，2007，(6)：3267～3272.

［4］ （美）Sunsan，Young Dave，Aitel 著，黑客防范手册. 吴世忠，郭涛，李斌，宋晓龙，等译. 北京：机械工业出版社，2005.

［5］ 胡建伟，汤建龙，杨绍全. 网络对抗原理. 西安：西安电子科技大学出版社，2004.

［6］ SPI Dynamics. Layer Seven：The Future of Vulnerabilities，2003.

［7］ 粟松涛，李春文，孙正顺. 一种新的 B/S 系统权限控制方法. 计算机工程与应用，2002，(1)：99～101.

［8］ Steven Splaine. Web 安全测试. 北京：机械工业出版社，2003.

［9］ Bret Hartman，Donald J. Flinn. 全面掌握 Web 服务安全性. 北京：清华大学出版社，2004.

［10］ 吴海翔，李宗伯. 让 Web 应用程序更安全. 2006 中国计算机网络安全应急年会暨亚太地区应急组织年会，2006.

［11］ 尹虹，李宗伯. Web 应用中的目录遍历漏洞的测试与实现. 通信与网络技术，2007.

［12］ Srivatsa，M. Iyengar，A. Jian Yin，Ling Liu. A Client-Transparent Approach to Defend Against Denial

of Service Attacks, Reliable Distributed Systems, 2006. SRDS 06. 25th IEEE SymPosium on oct. 2006, 61～70.

[13] Rubin, S. Jha, S. Miller, B. P. Automatic generation and analysis of NIDS attacks. Computer Security Applications Conference, 2004 (20): 28～38.

[14] Ismail, 0. Etoh, M. Kadobayashi, Y. Yamaguchi, S. A Proposal and implementation of automatic detection/collection system for cross-site scripting vulnerability, Advanced Information Networking and Applications. 2004. AINA 2004. 18th International Conference on Volume l, 2004, 145～151 Vol. 1

[15] Di Lucca, G. A. Fasolino, A. R. Mastoianni, M. Tramoniana, P. Identifying cross site scripting vulnerabilities in web applications. Web Site Evolution, 2004. WSE 2004. Proceedings. Sixth IEEE International Workshop on 11 Sept. 2004, 71～80

[16] (美)Greg Holden 著. 防火墙与网络安全——入侵检测和 VPNs. 王斌，孔璐译. 北京：清华大学出版社，2004.

[17] 魏红，动态网页技术 JSP 与 ASP、PHP 的比较浅析. 电脑知识与技术，2006，(2).

[18] 隋涛，基于 IIS 技术的 ASP 系统的安全问题. 情报与探索，2006，(11).

第 9 章　电子邮件安全

本章学习目标

随着 Internet 的快速发展和各种通信设施的不断完善,越来越多的敏感信息需要在 Internet 上传输。电子邮件是在 Internet 上传播信息的最主要的载体之一,其安全性越来越成为使用者和开发者关注的问题。本章主要介绍电子邮件安全的基本特性及其面临的安全问题,对电子邮件的几种安全技术进行全面介绍,包括 PGP、S/MIME、PEM、PKI 技术及安全防范措施。

通过对本章的学习,应掌握以下内容:

(1) 了解电子邮件安全的基本概念及电子邮件面临的安全问题。

(2) 掌握增强电子邮件安全性的几种安全技术。

(3) 掌握 PKI 技术。

(4) 掌握电子邮件安全防护措施。

电子邮件在日常生活中起着非常重要的作用,增强电子邮件的安全性是一项非常重要的任务。本章首先介绍电子邮件的基本概念、工作原理及特点,然后介绍电子邮件安全的概念及安全隐患、电子邮件几种常见的安全技术(PGP、S/MIME、PEM、PKI 等)、电子邮件的安全防范措施等内容。这些基本概念是读者全面熟悉与了解 E-mail 的开始,也是非常重要的基础知识。本章的很多内容将贯穿学习和使用 E-mail 的全过程,学习者通过本章的学习加深对电子邮件的了解,以便提高使用电子邮件的安全性。

9.1　电子邮件概述

9.1.1　电子邮件的基本概念

电子邮件就是人们通常所说的 E-mail。在英语中,mail 是邮件的意思,而 E 则是英语中电子一词 Electronic 的缩写。电子邮件是 Internet 上应用最广同时也是最基本的服务之一。只要能够连接到 Internet,拥有一个 E-mail 账户,就可以通过电子邮件系统,用非常低廉的价格、非常快的速度,与世界上任何一个角落的网络用户联络。

随着 Internet 的发展,电子邮件已经成为人们在网上互通信息的最常用的手段之一。通过电子邮件可以实现极为迅速的远距离通信,可以传输个人信息,或者向亲戚朋友致以问候,还可以传输语音、图像、视频等多媒体文件,也可以为电子商务服务。不论距离远近,完成整个过程只需要几分钟,价格也比普通的国际邮件便宜得多。此外,电子邮件还有一个显著的优点,就是无论身在何处,只要有一台能够连接 Internet 的计算机,就能随时随地收发电子邮件。

9.1.2 电子邮件的工作原理

电子邮件与普通邮件有类似的地方，发信者注明收件人的姓名与地址即邮件地址，发送方服务器把邮件传到收件方服务器，收件方服务器再把邮件发到收件人的邮箱中。图 9-1 所示为电子邮件的传输过程。

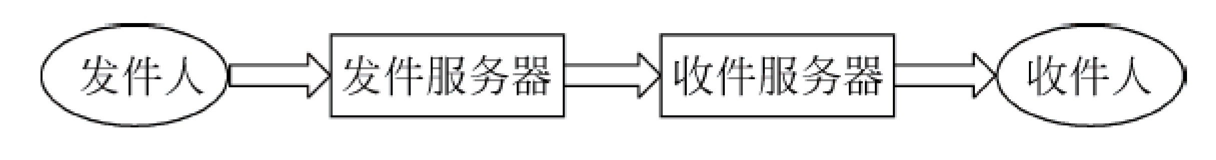


图 9-1 电子邮件传输过程

电子邮件的工作过程是：发信方通过客户端程序将编辑好了的电子邮件向服务器发送；发件服务器识别到收信人的地址，并向管理该地址的服务器发送消息；收件服务器识别后将消息存放在接收者的电子信箱内，并告知收信人有新邮件到来；收信人通过邮件客户程序连接到服务器后，就会看到服务器的通知，打开自己的电子信箱来查收邮件。

产生电子邮件安全隐患的原因主要有 3 个方面：

(1) 电子邮件传输协议自身的先天安全隐患。众所周知，电子邮件传输采用的是协议，它传输的数据没有经过任何加密，只要攻击者在其传输途中把它截获即可知道内容。

(2) 邮件接收客户端软件的设计缺陷。例如微软的客户端软件就存在可以使攻击者编制一些代码让木马或者病毒自动运行的安全隐患。

(3) 用户个人的使用问题。例如在网吧、学校等公共场所上网时把电子邮件的密码保存在上面，或者随意打开一些来历不明的邮件。

9.1.3 常见的电子邮件协议

1. SMTP（简单的邮件传输协议）

它是 Internet 上传输电子邮件的标准协议，用于提交和传输电子邮件，规定了主机之间传输电子邮件的标准交换格式以及邮件在链路层上的传输机制。SMTP 通常用于把邮件从客户端传输到服务器，或从某一服务器传输至另一服务器。

2. POP3（邮局协议）

POP3 是 Internet 上传输电子邮件的第一个标准协议，也是一个离线协议，它提供信息存储功能，负责为用户保存收到的电子邮件，并从邮件服务器上取回这些邮件。POP3 为客户端提供了发送信任状态（用户名和密码）的功能，规范了对电子邮件的访问。

3. IMAP4（交互式数据消息访问协议第 4 个版本）

当电子邮件客户机软件在笔记本计算机上运行时（通过慢速的电话线访问 Internet 和电子邮件），IMAP4 比 POP3 更常使用。使用 IMAP 时，用户可以有选择地下载电子邮件，甚至只是下载部分邮件，因此 IMAP4 比 POP3 更加复杂。

4. MIME（多功能的网际扩展协议）

Internet 上的 SMTP 传输机制是以 7 位二进制的 ASCII 码为基础的，适合传输文本文件，而声音、图像、中文等使用 8 位二进制编码的电子邮件，则需要进行 ASCII 转换（编码）才能在 Internet 上正确传输。MIME 增强了在 RFC 822 中定义的电子邮件报文的能力，允

许传输二进制数据。MIME 编码技术用于将数据从使用 8 位的格式转换为使用 7 位的 ASCII 码格式。

9.1.4　电子邮件的特点

对于大多数国内用户来说，收发 E-mail 是 Internet 上一个最常用的功能。为什么人们要用 E-mail 收发邮件呢？因为它和普通邮件相比有很多优点。

（1）方便。E-mail 非常方便，足不出户就可以和远在万里之外的其他人通信。而且用户的信箱与普通信箱不同，是存在于 Internet 上的电子信箱，只要能连接到 Internet 就能随时随地读取和发送邮件。此外，还能把同一封信同时发给好几个不同的朋友。

（2）快捷。Internet 上的信息在光纤中是以光速传播的，因此 E-mail 比普通的邮政信件快得多，甚至比普通的电报还要快。在网络通畅的情况下，一封几 KB 的 E-mail 邮件只要几分钟就能到达收信人的电子信箱，不论其信箱是在国内还是国外。

（3）便宜。对于拨号上网的用户，为了尽量节省上网费用，通常应该在没有联网的时候把信写好。由于收发 E-mail 所占用的上网时间很短，所以相对寄送普通邮件来说，E-mail 是很便宜的，尤其对收发国际邮件的用户更是如此。

（4）信息多样。寄送普通信件，信息的量和种类十分有限。E-mail 则不同，它能把可以用数字表示的所有信息以附加文件的方式发给收件人，这些信息可以是文字、图像，也可以是声音甚至动画等形式的多媒体文件。

（5）功能强大。E-mail 不仅可以用来向网上的亲朋好友发邮件，还可以参加范围广泛的专题讨论组（Mailing List）、订阅电子期刊、完成文件传输（FTP）等功能。

9.2　电子邮件安全概述

目前国内和国际上使用的电子邮件一般都是免费电子邮件，除了简单的密码认证以外没有其他任何安全措施，用户的所有信息都被暴露在攻击者面前，这在很大程度上阻碍了网上各种业务的进一步发展。因为很少有人愿意在没有任何保障的情况下把一些重要信息放到网络上传输。即使目前有几个邮件服务提供商声称自己提供邮件保密业务，他们也只是在邮件传输的某几个环节中保证邮件内容不被泄露，并不能从每个环节上保证信息不被泄露。他们的产品更不能提供身份认证、防抵赖以及防篡改等安全服务。

网络信息安全是一个综合、交叉的学科领域，它要利用数学、电子、信息、通信、计算机等诸多学科的长期知识积累和最新的发展成果。网络信息安全研究的内容很多，它涉及安全协议、密码理论、信息分析等。其中密码是网络信息安全的关键技术。目前网络信息安全的目标领域主要由两部分组成：网络信息交换安全和电子商务信息安全。电子邮件安全属于信息安全的一部分，它是信息安全中最具代表性的一种，对它的研究可以推广到网络信息安全的多个方面，尤其是对电子商务的安全具有很大的参考意义，无论是以后的推广，还是安全邮件本身都有很大的使用价值。

电子邮件目前面临的主要安全问题有：

1. 邮件病毒

邮件病毒一般是通过邮件中的附件进行扩散的,一旦运行了附件中的病毒程序,就能导致计算机染毒。然而一些新型邮件病毒只需在打开邮件正文或浏览标题时就能够感染计算机。邮件已成为计算机病毒传播的一个主要途径。目前多数蠕虫病毒都可以通过邮件传播。邮件蠕虫病毒将病毒邮件发送给搜索到的邮件地址,一旦用户打开带有病毒的邮件或运行病毒程序,该计算机就会马上感染病毒。蠕虫病毒能在感染的系统中收集邮件地址,并发送大量病毒邮件。一些邮件服务器每小时能收到上万封病毒邮件,有时甚至会造成网络堵塞或邮件服务器瘫痪。还有些蠕虫病毒能利用人工漏洞传染病毒。

2. 垃圾邮件

中国互联网协会在《中国互联网协会反垃圾邮件规范》中定义的垃圾邮件包括如下属性:

(1) 收件人事先没有提出要求或者同意接收的广告、电子刊物、各种形式的宣传品等宣传性的电子邮件。

(2) 收件人无法拒收的电子邮件。

(3) 隐藏发件人身份、地址、标题等信息的电子邮件。

(4) 含有虚假的信息源、发件人、路由等信息的电子邮件。

垃圾邮件不仅占用了大量的网络资源,而且浪费了邮件服务使用者宝贵的时间和精力。2007 年中国网民收到的垃圾邮件总量为 694 亿封,人均收到垃圾邮件高达 330 封,造成约为 187.2 亿元人民币损失。虽然采用了各种技术手段,赛门铁克统计仍显示:2009 年 6 月,垃圾邮件量依然约占所有电子邮件信息量的 90%。

3. 监听

监听可分为两种方式:

(1) 局域网内的监听。通常使用嗅探器对局域网内传输的数据进行监听,由于协议通常都是明文传输,嗅探器很容易嗅探到用户的邮箱密码。因此使用浏览器收发邮件就显得安全一些。

(2) 来自邮箱内部的监听。用户密码被破解之后,攻击者并不会修改密码,而是将邮件先发送到攻击者的信箱,再将邮件转发到用户邮箱,从而完全控制用户信箱的流量,选择其能够接收到的邮件。这种监听方法相当隐蔽,危害很大。

9.3 几种电子邮件安全技术

由前面的内容可知,电子邮件服务存在的主要安全问题包括 3 个方面:敏感信息泄露、病毒传播和垃圾邮件泛滥。这些问题已经成为阻碍信息系统电子化进程的重要因素,要求安全电子邮件的呼声也越来越高,安全电子邮件的标准也在不断完善。

9.3.1 PGP

PGP(Pretty Good Privacy)是由 Phil Zimmermann 主要开发的网络应用,能为电子邮

件系统和文件存储应用过程提供认证业务和保密业务。Phil Zimmermann 所做的工作主要有：

(1) 选择了最好的加密算法来创建数据块。

(2) 将密码算法集成在与操作系统和处理器独立且其命令集易于使用的应用程序中。

(3) 能使软件包或源代码等文档通过 Internet、公告板或其他商用网络免费使用。

(4) 用户可与公司建立协议，以获得完全兼容的、低成本的商用版本。

PGP 被广泛使用的原因有：

(1) 世界各地都可以免费使用的几种版本均可用于各种平台，包括 DOS、Windows、UNIX 和 Macintosh。此外，商用版支持商家开发自己的产品。

(2) 所用的算法具有很高的安全性，其软件包中的公钥加密算法有 RSA、DSS 以及 ElGamal，单钥算法有 CAST_128、IDEA、三重 DES 和 Hash 算法 SHA。

(3) 适用范围极为广泛，从公司到个人都可以使用，公司可用它作为加密的标准方案，个人可用它和世界各地安全通信。

(4) 它的开发未受任何政府组织和标准化组织的控制。

下面详细介绍 PGP 的运行方式、密钥的产生和存储以及公钥的管理。

1. 运行方式

PGP 的 5 种业务包括认证性、保密性、压缩、电子邮件的兼容性、分段。表 9-1 总结了这 5 种业务。其中 CAST-128 是由加拿大 Carlisle Adams 与 Stafford Tavares 设计的分组密码，已在 RFC2144 中公布。算法具有传统的 Feistel 网络结构，采用 16 轮迭代，明文分组长度为 64 比特，密钥长度以 8 比特为增量，从 40 比特到 128 比特可变。

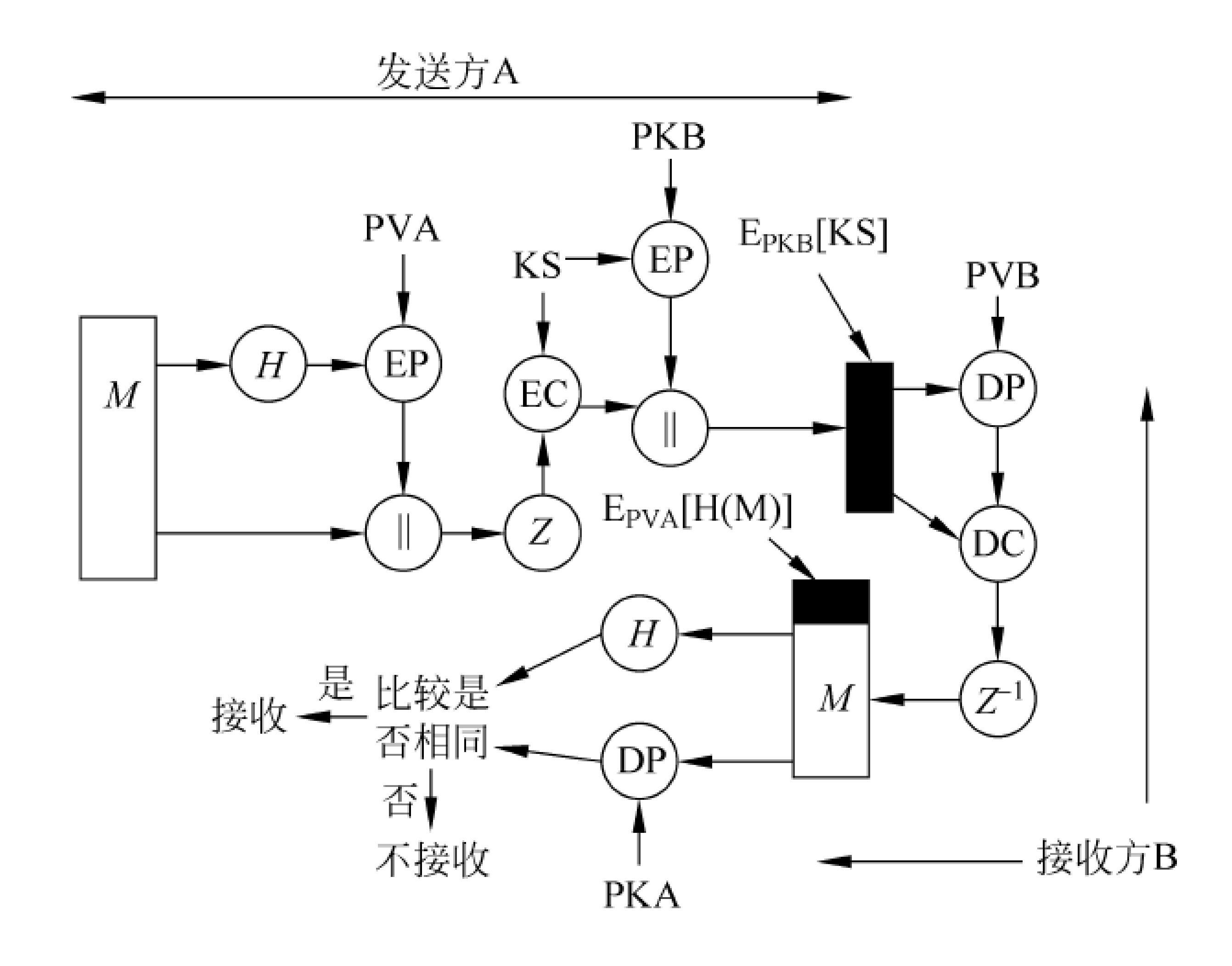
表 9-1　PGP 的业务

功　能	所 用 算 法	描　述
数字签名	SS/SHA 或 RSA/SHA	发送方使用 SHA 产生消息摘要，再用自己的密钥按 DSS 对消息摘要签名
消息加密	AST 或 IDEA 或 3 个密钥的三重 ES/E1Gamal 或 RSA	消息由用户产生的一次性会话密钥按 CAST-128 或 IDEA 或三重 DES 加密，用接收方的公钥按 ElGamal 或 RSA 加密会话密钥
压缩	ZIP	消息经 ZIP 算法压缩后存储或发送
电子邮件的兼容性	Base64 编码	使用 Base64 编码将加密的消息转换为 ASCII 字符串，提供了电子邮件应用系统透明性
分段		对消息进行分段和重组以适应 PGP 对消息最大长度的限制

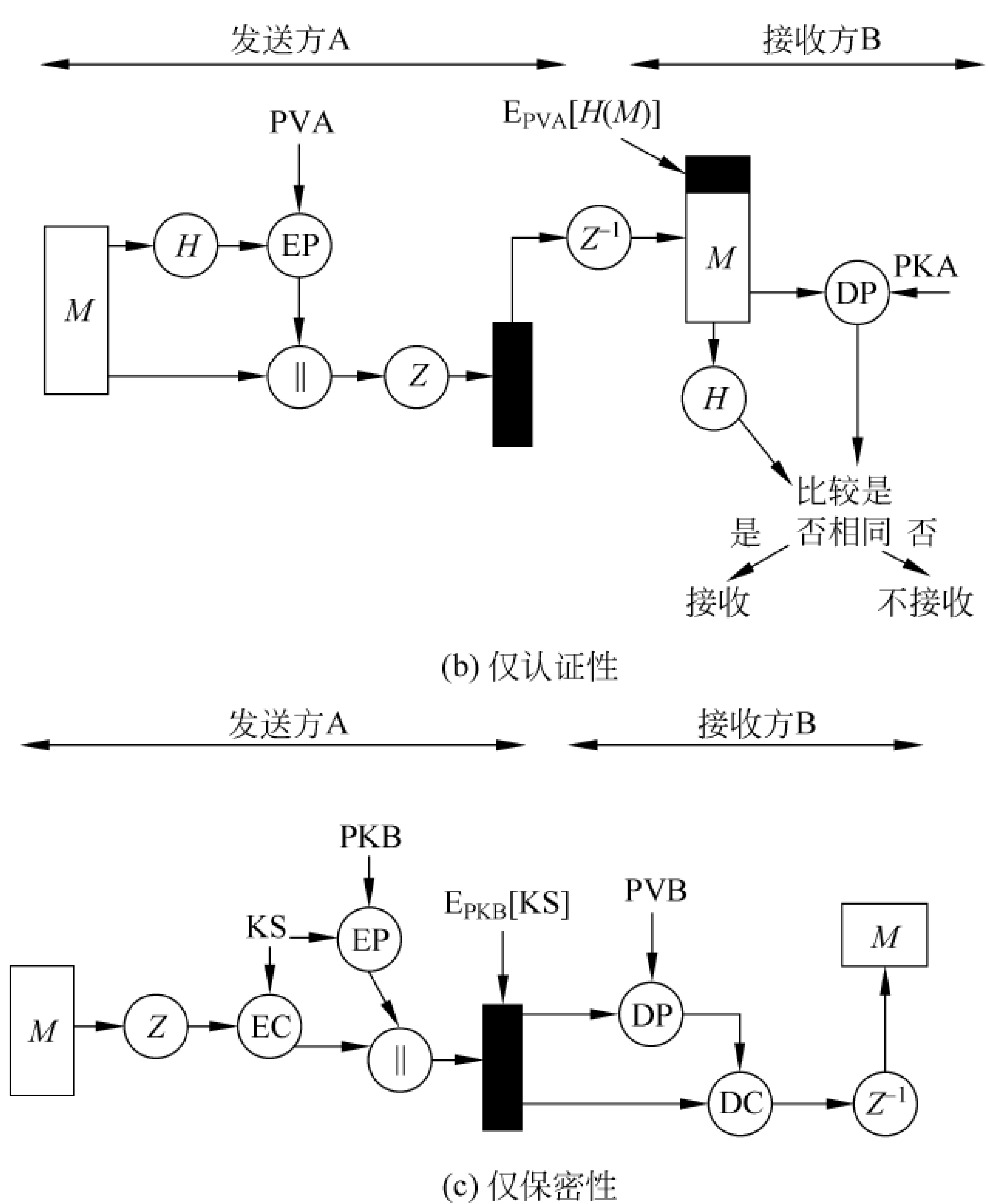
图 9-2 所示为 PGP 的认证业务和保密业务示意图，其中 KS 为分组加密算法所用的会话密钥，EC 和 DC 分别为分组加密算法和解密算法，EP 和 DP 分别为公钥加密算法和解密算法，PVA 和 PKA 分别为用户 A 的私钥和公钥，PVB 和 PKB 分别为用户 B 的私钥和公钥，H 表示 Hash 函数，‖ 表示连接，Z 表示 ZIP 压缩算法，M 表示消息本身。

1) 认证业务

PGP 中通过数字签名提供认证的过程分为 5 步：



(a) 既有认证性又有保密性



(b) 仅认证性

(c) 仅保密性

图 9-2 PGP 的认证业务和保密业务示意图

(1) 发送方产生消息 M。

(2) 用 SHA 产生 160 比特的消息摘要 $H(M)$。

(3) 发送方用自己的密钥 SKA 按 RSA 算法对 $H(M)$进行加密,并将加密结果与 M 连接在一起发送。

(4) 收送方用发送方的公钥对得到的消息摘要进行解密得到 $H(M)$。

(5) 接收方对收到的 M 计算消息摘要，并与(4)中的 $H(M)$ 比较，如果一致则认为 M 是真实的。

过程中结合使用了 SHA 和 RSA 算法，类似的也可结合使用 DSS 和 SHA 算法。以上过程将消息的签名与消息连接在一起发送或存储，但在有些情况中将消息的签名与消息分开发送或存储(称为干净签名)。例如将可执行程序的签名分开存储，可用来检查程序是否被病毒感染。再例如多人签署同一文件(例如法律合同)，每人的签名都应与被签文件分开存储，否则第一个人签完字后将消息与签名连接在一起，第二个人签名是既要签消息，又要签第一个人的签名，从而会形成签名的嵌套。

2) 保密业务

PGP 的另一业务是为传输或存储的文件提供加密的保密性业务。加密算法可用 CAST-128、IDEA 或三重 DES，运行模式为 64 比特 CFB 模式。加密算法的密钥为一次性的，即每加密一次消息都需产生新的密钥，称为一次性会话密钥。新密钥也要用接收方的公钥加密后与消息一起发往接收方，整个过程如下：

(1) 发送方产生消息 M 及一次性会话密钥 KS。

(2) 用密钥 KS 按 CAST-128/IDEA/3DES 加密 M。

(3) 用接收方的公钥 PKB 按 RSA 算法加密一次性会话密钥 KS，将(2)、(3)中的两个加密结果连接起来一起发往接收方。

(4) 接收方用自己的私钥按 RSA 算法恢复一次性会话密钥。

(5) 接收方用一次性会话密钥恢复收到的消息。

PGP 为加密一次性会话密钥还提供了 E1Gamal 算法以供选用。

以上方案有以下几个优点：

(1) 由于分组加密速度远快于公钥加密算法速度，因此使用分组加密算法加密消息、公钥加密算法加密一次性会话密钥能大大地减少加密时间。

(2) 由于会话密钥是一次性的，因此没有必要使用会话密钥的交换协议。同时，由于电子邮件的存储转发特性，也无法使用握手交换协议。本方案使用公钥加密算法来传输一次性会话密钥，保证了仅接收方能得到邮件。

(3) 一次性会话密钥的使用进一步加强了本来就很强的分组加密算法，因此只要公钥加密算法是安全的，整个方案就是安全的。

PGP 允许用户可选择的密钥长度范围为 768～3072 比特，而若是使用 DSS，其密钥限制为固定的 1024 比特。

3) 保密性与认证性

如果对同一消息同时提供保密性和认证性，可使用图 9-2(a)的方式。发送方首先用自己的私钥对消息签名，将明文消息和签名连接在一起，再使用一次性会话密钥按 CAST-128/IDEA/3DES 对其加密，同时用 RSA 对会话密钥加密，最后将两个加密结果一同发往接收方。这一过程中，先对消息签名再对签名进行加密。这一顺序优于先加密、再对加密结果签名。因为将签名与明文消息一起存储会带来很多方便，同时也给第三方签名的认证带来方便。

4) 压缩

压缩的目的是为邮件的传输或文件的存储节约空间。压缩运算应在签名之后、加密以

前,这是因为:

(1) 压缩之前要完成签名有两个原因:对不压缩的消息签名,可便于以后对签名的验证。如果对压缩后的消息签名,在要对签名进行验证时,需存储压缩后的消息或在验证签名时对消息重做压缩;即使用户愿意对压缩后的消息签名且愿意验证时对原消息重做压缩,实现起来也极为困难,这是由于ZIP压缩算法具有不确定性的,该算法在不同的实现中会在运行速度和压缩率之间进行不同的折中,从而产生不同的压缩结果(虽然解压结果相同)。

(2) 对消息压缩后再进行加密可加强其安全性,这是因为消息压缩后比压缩前的冗余度要小,因此会使密码分析更为困难。

5) 电子邮件的兼容性

PGP以上3种业务中,传输的消息都有被加密的部分(或所有部分),这些部分构成了任意8比特位组串。然而许多电子邮件系统只允许使用ASCII文本串,为此PGP提供了将8比特位串转换为可打印的ASCII字符的服务。转换方法是Base64编码,将每3个8比特位组的二元数据映射为4个ASCII字符。Base64编码可将被变换的消息扩张33%,但是由于扩展的是会话密钥和消息的前部分,而这一部分又是比较紧凑的,因此对明文消息的压缩足以弥补Base64编码引起的扩展。有实例显示,ZIP的平均压缩率大约为2.0。因此如果不考虑相对小的签名和密钥部分,对长度为x的文件来说,压缩和扩展的总体效果为$1.33\times 0.5\times x=0.665x$,即总体上有三分之一的压缩。

PGP变换具有"盲目性",即不管输入变换的消息内容是不是ASCII文件,都将变换为Base64格式。因此在图9-2(b)所示的仅提供认证的服务中,对消息及其签名进行Base64编码,编码后的结果对不经意的观察者是不可读的,从而能提供一定程度的保密性。作为一种配置选择,PGP可以只将消息的签名部分转换为Base64格式,从而使得接收方不使用PGP就可以阅读消息,但对签名的验证仍然需要使用PGP。

图9-3所示为PGP的消息处理过程示意图,其中(a)图和(b)图分别是发送方和接收方对消息的处理过程示意图。发送方首先对消息的Hash值签名(如果需要的话),然后明文消息及其签名(如果有的话),再经过压缩函数进行压缩。如果要求保密性,则用一次性会话密钥按分组加密算法加密压缩结果,同时用公钥加密算法加密一次性会话密钥。将两个加密结果连接在一起后,再经Base64编码变换成Base64格式。

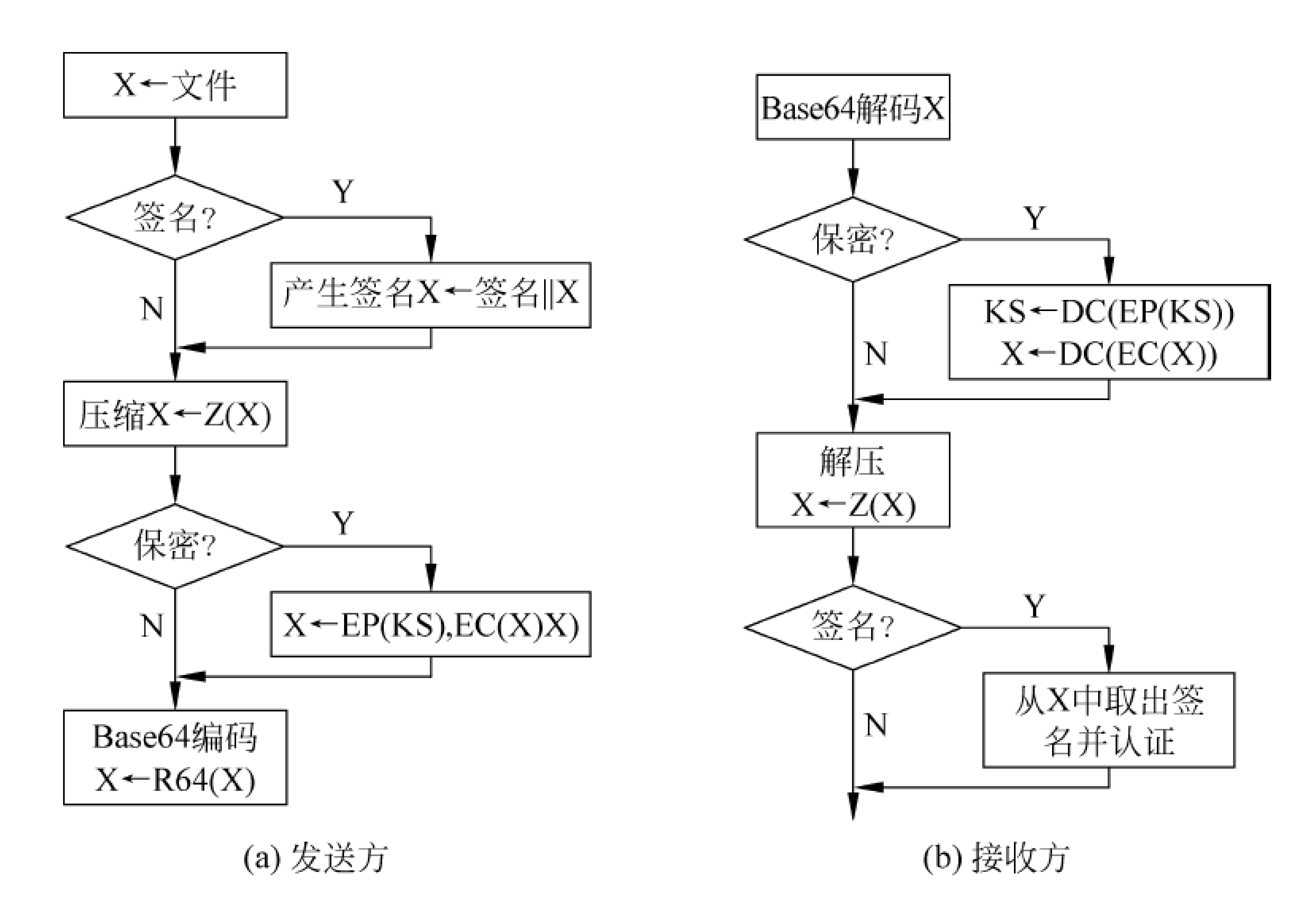
接收方首先将接收到的结果进行Base64解码。接着,如果消息是加密文,则恢复一次性会话密钥,由一次性会话密钥恢复加密的消息,并对之解压。如果消息还经过签名,则从上一步恢复的消息中取出消息的Hash值,并与自己计算的消息的Hash值进行比较。

6) 分段与重组

电子邮件通常都对最大可用的消息长度有所限制,如果消息长度大于最大可用长度,则将消息分为若干子段并分别发送。分段操作在图9-3所示的Base64变换以后进行,会话密钥报头和签名报头仅在第一子段的开头处出现一次。接收方在图9-3(b)的处理过程之前,先去掉第一子段开头处的报头再将各子段拼接在一起。

2. 密钥和密钥环

PGP所用的密钥有4类:分组加密算法所用的一次性会话密钥、基于密码短语的密钥、公钥加密算法所用的公钥和私钥。PGP必须满足以下3个要求:



(a) 发送方　(b) 接收方

图 9-3　PGP 的消息处理过程示意图

（1）能够产生不可猜测的会话密钥。

（2）用户可有多个公/私钥对。这是因为一方面用户可能希望随时更换自己的密钥对，另一方面用户可能希望在同一时间和多个通信方同时通信时分别使用不同的密钥对，或者用户可能希望通过限制一个密钥加密内容的数量来增加安全性。因此用户和它的密钥对不是一一对应的关系，必须采取某一方式对密钥加以识别。

（3）PGP 的每一个用户都必须对存储自己密钥对的文件加以维护，同时还需对存储所有通信双方公钥的文件加以维护。

1）会话密钥的产生

会话密钥的使用是一次性的，其中 CAST-128 和 IDEA 所用会话密钥长度为 128 比特，3DES 所用的会话密钥长度为 168 比特。下面以 CAST-128 为例来介绍其密钥的生成。

产生 CAST-128 密钥的随机序列产生器由 CAST-128 加密算法构成。其输入为一个 128 比特的密钥和两个 64 比特的明文，采用 CFB 模式，对两个明文分组加密，再将得到的两个 64 比特密文分组连接在一起形成所要产生的 128 比特密钥。其中初始的两个 64 比特的明文分组由用户随机的键盘输入得到，将输入的一个字符表示成 8 比特的数值，共随机输入 12 个字符，得到 96 比特的数值，剩下 32 比特则用来表示键盘输入所用的时间。随机序列产生器输入的 128 比特的密钥则取它上一次输出的 128 比特的会话密钥。

2）密钥识别符。

如前所述，PGP 在对消息加密的同时，还需用接收方的公钥加密一次性会话密钥，从而使得只有接收方能恢复会话密钥，进而恢复加密的消息。如果接收方只有一个密钥对（即公钥/私钥对），即可直接恢复会话密钥。然而接收方通常都有多个密钥对，如何知道会话密钥是用哪一个公钥加密的呢？一种解决办法是发送方将所用的接收方的公钥与邮件一起发送给接收方，但这种方法对空间的浪费太多，RSA 的公钥长度可达数百位十进制数。另一种办法是对每一个用户的每一个公钥都指定一个唯一的识别符，称为密钥 ID，发送方将所使用的公钥的 ID 发给接收方。但是这种情况下必须考虑密钥 ID 的存储和管理，并且收发双

方都必须能够从密钥 ID 得到对应的公钥，这造成了不必要的负担。PGP 采用的方式是使用公钥中 64 个最低有效位表示该密钥的 ID，即公钥 PKA 的 ID 是 PKA mod 2 的 64 次方。由于 64 位足够长，因此不同密钥的 ID 相重的概率很小。PGP 在数字签名时也将密钥加上识别符，这是因为发送方签名时可能有很多私钥可供使用，接收方必须知道使用发送方的哪个公钥来验证签名。PGP 用签名中的 64 比特来表示相应公钥的 ID。

3）PGP 的消息格式

图 9-4 所示为 PGP 中发送方 A 发往接收方 B 的消息格式，其中 EPKB 表示用接收方 B 的公钥加密，ESKA 表示用发送方 A 的私钥加密（即 A 的签名），ZIP 是压缩算法，EKS 表示用一次性会话密钥 KS 加密，R64 是 Base64 编码。

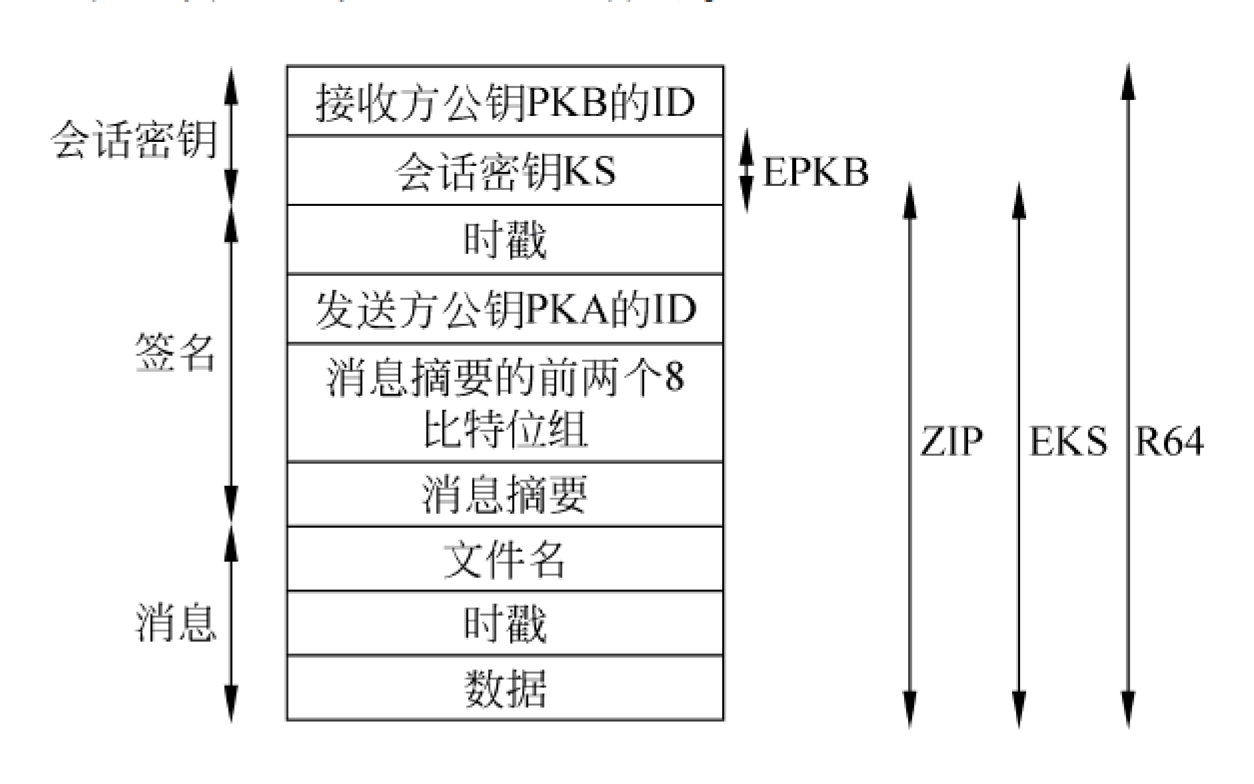


图 9-4　PGP 的消息格式

PGP 的消息由 3 个部分组成：消息、消息的签名（可选）、会话密钥（可选）。

消息部分包括被存储或被发送的实际数据、文件名以及时戳，时戳表示产生消息的时间。

签名部分包括以下成分：

（1）时戳：产生签名的时间。

（2）消息摘要：消息摘要是 SHA 对消息签名的时戳与消息本身连接得到 160 比特输出后再由发送方用私钥签名的结果。消息摘要中包含签名时戳的目的是防止重放攻击，而不包含消息的文件名和产生消息的时戳的目的是使得分离的签名与作为前缀而附加在消息前的签名完全一样。分离的签名是由无报头域的实际数据计算得到的。

（3）消息摘要的前两个 8 比特位组：接收方利用解密消息摘要后得到的前两个 8 比特位组来确定自己在验证发送方的数字签名时是否正确地使用了发送方的公钥。消息摘要的前两个 8 比特位组也可用作消息的 16 比特帧校验序列。

（4）发送方公钥的 ID：用于表示解密消息摘要（即验证签名）的公钥，相应地也代表了签名的私钥。

消息部分和签名部分经 ZIP 算法压缩后再用会话密钥加密。会话密钥部分包括会话密钥和接收方公钥的标识符，标识符用来识别发送方加密会话密钥使用的是接收方的哪一个公钥。在发送消息前，对整个消息做 Base64 编码。

4）密钥环

为了有效存储、组织密钥，同时也为了便于用户使用，PGP 为每个节点（即用户）都提供

了两个表型的数据结构,一个用于存储用户自己的密钥对(即公私钥对),另一个用于存储该用户所知道的其他各用户的公钥。这两个数据结构分别称为私钥环和公钥环,如表 9-2 所示。

表 9-2(a)　密钥环——私钥环

时戳	公钥 ID	公钥	被加密的私钥	用户 ID
Ti	PKi mod 2 的 64 次方	PKi	PVi	USERi

表 9-2(b)　密钥环——公钥环

时戳	密钥 ID	公钥	拥有者可信字段	用户 ID	密钥合法性字段	签名	签名可信字段
Ti	PKi mod 2 的 64 次方	PKi	Trust flagi	USERi	Trust flagi		

在私钥环中,每行表示该用户的一个密钥对,其数据项有:产生密钥对的时戳、密钥 ID、公钥、被加密的私钥和用户 ID,其中公钥 ID 和用户 ID 可作为该行的标识符。用户 ID 可用用户的邮件地址来表示,可以为一个密钥对使用多个不同的用户 ID,也可在不同的密钥对中使用相同的用户 ID。

私钥环由用户自己存储,仅供用户自己使用,而且为了使私钥尽可能地安全,私钥会通过 CAST-128(或 IDEA 或 3DES)加密后以密文的形式存储。加密过程为:用户首先选择一个密码作为 SHA 的输入,产生一个 160 比特的 Hash 值后销毁密码,再用 Hash 值的 128 比特作为密钥按 CAST-128 对私钥加密,加密完成后再销毁 Hash 值。以后若要取出私钥,必须重新输入密码,PGP 产生出密码的 Hash 值,并以此 Hash 值为私钥按 CAST-128 解密被加密的私钥。

由加密私钥的过程可见,私钥的安全性取决于所用密码的安全性,因此用户使用的密码应该是易于记忆的但又是不易被他人猜出的。

公钥环中存储的是该用户所知道的其他用户的公钥,其数据项包括:时戳、密钥 ID、公钥、用户 ID,其中密钥 ID 和用户 ID 可作为该行的标识符。

下面介绍消息传输和接收时密钥环是如何使用的。为简单起见,下面过程省略了压缩过程和 Base64 转换过程。假定消息既要被签名又要被加密,则发送方 A 需执行以下过程(如图 9-5 所示,其中 RNG 表示随机序列发生器,其他符号和图 9-2 中的相同):

(1) 签署消息:

① PGP 使用 A 的用户 ID 作为索引(即关键字)从 A 的私钥环中取出 A 的私钥。如果用户 ID 位为默认值,则从私钥环中取出第一个私钥。

② PGP 提示用户输入密码用于恢复被加密的私钥。

③ 由 A 的私钥产生消息签名。

(2)加密消息:

① PGP 产生一个会话密钥,并由会话密钥对消息及签名加密。

② PGP 使用接收方 B 的用户 ID 作为关键字,从公钥环中取出 B 的公钥。

③ PGP 用 B 的公钥加密会话密钥形成发送消息中的会话密钥部分。

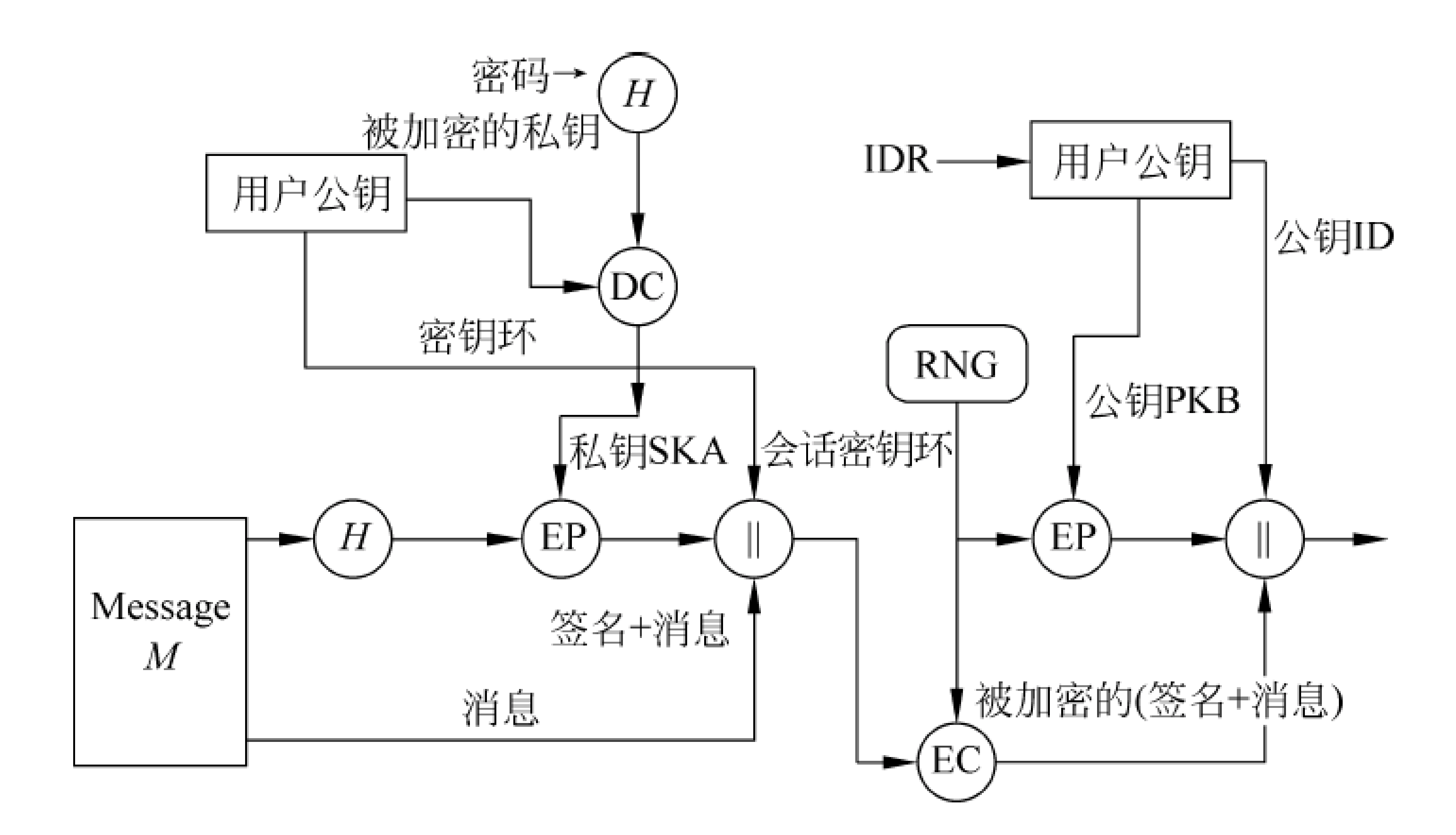


图 9-5 PGP 的消息产生过程

PGP 的消息接收过程如图 9-6 所示。

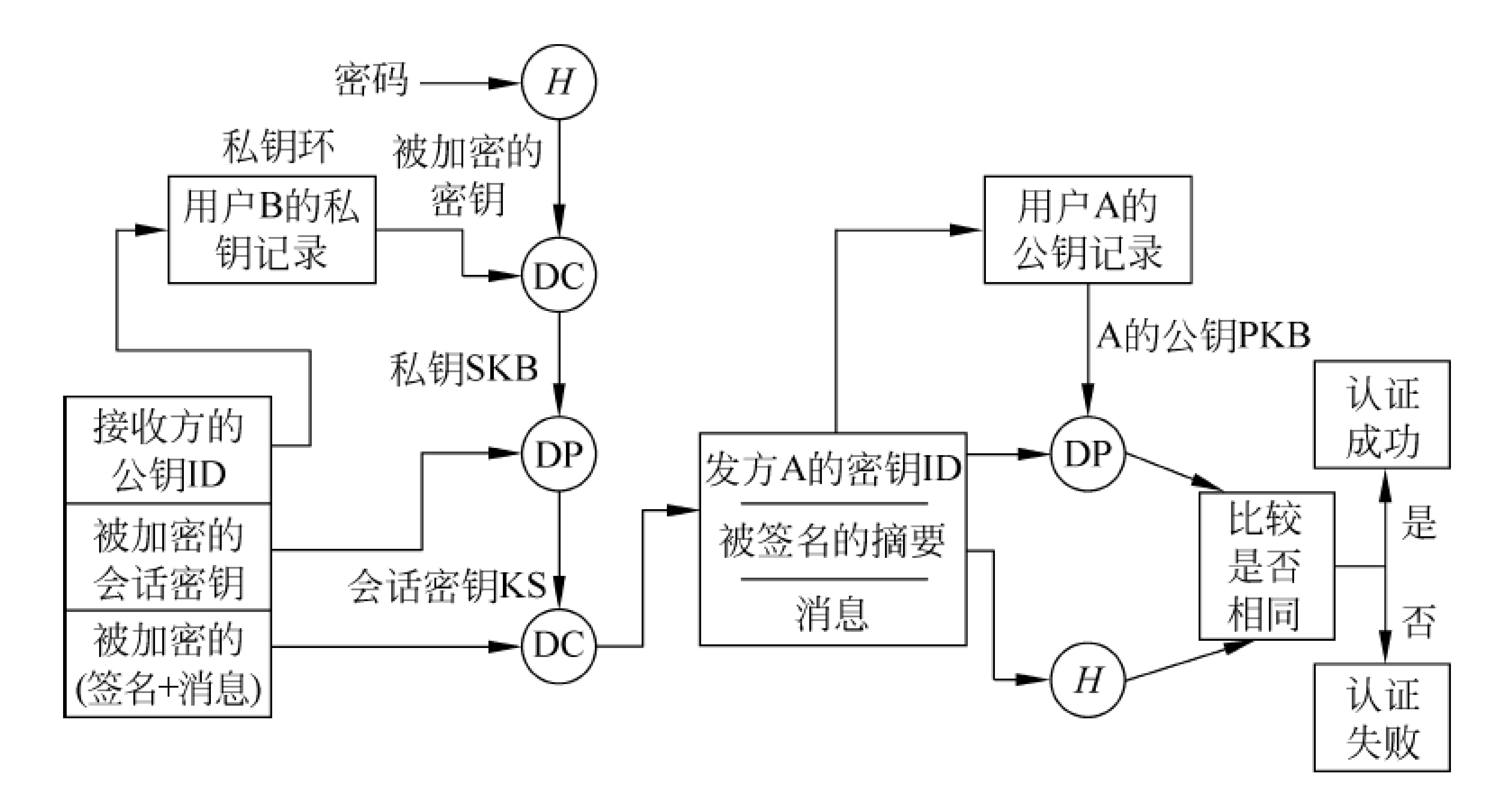


图 9-6 PGP 的消息接收过程

(1) 解密消息：

① PGP 从接收到的消息的会话密钥部分取出接收方 B 的密钥 ID,并以此作为关键字从 B 的私钥环中取出相应的被加密的私钥。

② PGP 提示 B 输入密码以恢复私钥。

③ PGP 用私钥恢复出会话密钥,并进而解密消息。

(2) 认证消息：

① PGP 从收到的消息的签名部分取出发送方 A 的密钥 ID,并以此作为关键字从发送方的公钥环中取出发送方的公钥。

② PGP 用发送方的公钥恢复消息摘要。

③ 对收到的消息重新计算消息摘要,并对恢复出的消息摘要进行比较。

9.3.2 S/MIME

S/MIME(Secure/Multipurpose Internet Mail Extension,安全/多用途 Internet 邮件扩

展)是基于 RSA 数据安全技术、对 MIME Internet 邮件格式标准的安全增强版。虽然 PGP 和 S/MIME 都是沿袭 IETF 标准,但 S/MIME 可能会作为商用的行业标准,而 PGP 则仍然用于个人目的的电子邮件安全。

1. S/MIME 的安全功能

S/MIME 的安全功能与 PGP 的安全功能相同,有加密和签名,即:

(1) 数据的封装:被封装的数据包括被加密的消息内容和被加密的消息加密密钥,封装过程与 PGP 类似,只是没有压缩过程。

(2) 数据的签名:发送方用自己的私钥对消息摘要进行签名,然后将签名和消息内容一起经 Base64 编码后发给对方,仅当接收方也具有 S/MIME 的安全功能时才能解读该消息。或者将邮件内容和签名分开发送(又称干净签名),这种情况下,接收方即使没有 S/MIME 功能也能够阅读邮件,只是无法对其签名进行认证。

提示:签名分为干净签名和模糊签名,在干净签名中消息与签名分开存放,即使接收方不支持 S/MIME 也能看到邮件内容;模糊签名将签名与消息混在一起,如果接收方不支持 S/MIME 将看到一些乱码,但它并不是加密,了解其格式,经过变换后仍能看到邮件内容。

安全功能的另一种方式是将加密和签名嵌套使用,即加密后的数据又可被签名,或者先签名后加密。

2. S/MIME 使用的密码算法

S/MIME 使用的密码算法有消息摘要算法、签名算法、加密会话密钥的公钥加密算法和加密消息的单钥加密算法。建议使用的消息摘要算法是 SHA(但也支持 MD5)。签名的首选算法是 DSS,加密会话密钥的首选算法是 ElGamal。此外,RSA 算法也能用于签名和加密会话密钥。加密消息的建议算法是 3DES,也支持 40 比特的 RC2,后者虽然强度弱,但不受美国出口限制,前者美国禁止出口。RC2 是 R. Rivest 设计的分组加密算法,其明文分组和密文分组都为 64 比特,密钥长度在 8～1024 比特可变。

3. S/MIME 的消息格式

S/MIME 在 MIME 消息格式的基础上增加了签名、加密等安全功能,在消息报头中新增了两个内容类型:multipart 和 application,并在 application 的子类型中标有 PKCS (Public-Key Cryptography Specifications)。PKCS 是 RSA 实验室发布的公钥密码规范。

S/MIME 的消息产生过程如下:首先按 MIME 消息产生的规范形式产生 MIME 消息,再加上与安全有关的数据,例如算法标识符、证书等。然后由 S/MIME 对上述结果进行处理(包括加密和签名)生成自己的消息内容(称为 PKCS 对象),最后加上 MIME 报头就形成了完整的 S/MIME 邮件消息。由于加密、签名的作用,邮件消息(或部分)形式是二元数字序列,还需对其做 Base64 编码转换变成规范形式。

下面分别介绍 S/MIME 新增加的两个内容类型。

1) multipart

该类型又有一个子类型 signed,表示发送的内容是明文消息和对消息的签名。

2) application

(1) application/pkcs7-mine/EnvelopedData:pkcs7-mine 表示子类型;Enveloped Data

是类型参数，用于表示 S/MIME 对邮件消息的 Base64 编码方式，后面介绍的其他子类型中的参数也表示编码变换方式。

EnvelopedData 表示的编码方式是对消息进行封装，其封装过程如下：

① 产生分组加密算法(例如 3DES 或 Rc2/40)所需的会话密钥。

② 用接收方的公钥加密会话密钥。

③ 为接收方产生一个数据包，包括 X.509 公钥证书、加密会话密钥所用算法的标识符、加密的会话密钥。

④ 用会话密钥加密消息内容。

第③步产生的数据包和第④步产生的加密内容一起构成了封装的消息，再对封装的消息进行 Base64 编码转换。接收方对收到的内容首先做 Base64 解码变换，然后用自己的私钥恢复出会话密钥，最后再用会话密钥对加密的消息内容解密。

(2) application/pkcs7-mine/signdData：参数 signedData 表示对消息签名，过程如下：

① 由消息摘要算法 SHA 或 MD5 求得消息被签名的摘要值。

② 用签名者的私钥对消息摘要签名。

③ 为接收方产生一个数据包，包括签名者的证书、消息摘要算法标识符、签名算法标识符和签名结果。

消息的签名可能由多个签名者产生，因此最后得到的签名内容包括：消息本身、消息摘要算法标识符、第③步产生的若干数据包，还可能包括可构造出由顶层证书发放机构到签名者的证书链的一组公钥证书。

签名内容经 Base64 编码转换后发往接收方，接收方对收到的内容先做 Base64 解码变换，再用发送方的公钥对签名解密得到消息摘要。同时由收到的内容计算出消息摘要，对两个摘要进行比较以验证发送方的签名。

(3) applicafion/pkcs7-signature：表示类型为 multipart/signed 的消息中签名部分的内容类型。multipart/signed 类型的邮件消息是由明文消息和对消息的签名两部分组成的，其中第一部分的产生方式应保证它在从发送方发往接收方的整个过程中不被篡改。如果第一部分的字符不是 7 比特的 ASCII 字符，就需对其进行 Base64 编码变换或 Quoted-printable 编码变换，变换方式与上述 application/pkcs7-mine/signedData 的变换方式相同。然后再对第二部分即消息的签名进行 Base64 编码变换，得到的类型就是 application/pkcs7-signature。

(4) application/pkcs7-mine：表示用户或应用程序将向证书发放机构发送一组数据，用来申请一个公钥证书。这一组数据包括证书请求信息、公钥加密算法标识符和申请者对证书请求信息的签名。证书请求信息又包括证书主题名称和用户公钥的比特串表示。

(5) application/pkcs7-mine/degenerate signedData：表示该消息是证书发放机构对用户申请公钥证书请求的应答，消息的内容只有发给用户的证书或证书吊销列表。

9.3.3 PEM

PEM(Privacy-Enhanced Mail)是基于 X.509v1 提出的一个专用于加密 E-mail 通信的正式 Internet 标准。RFC1421、RFC1422、RFC1423 和 RFC1424 规定的 Internet 保密增强邮件标准 PEM 在邮件的保密安全性方面大大强于 SMTP 和 POP3 协议。RFC1421 介绍

了消息加密和验证过程,RFC1422 给出了基于证书的密钥管理,RFC1423 描述了算法、模式和身份证,RFC1424 描述了密钥证书和相关服务。

PEM 使用两级密钥:数据加密密钥(DEK)和交换密钥(IK)。DEK 用来加密消息正文和计算消息集成校验(MIC),同时用来加密 MIC 的签名表示。DEK 在每次会话中都会重新生成一个,从而达到一次一密的效果。而 IK 用来加密 DEK,以便在每次会话的初始段对 DEK 进行加密交换。加密 DEK 的 IK 是接收方的公钥。加密 MIC 的 IK 是发送方的私钥,实现对 MIC 的签名。依据 PEM 的机制,邮件加密签名生成过程、邮件接收和验证过程分别如图 9-7 和图 9-8 所示。

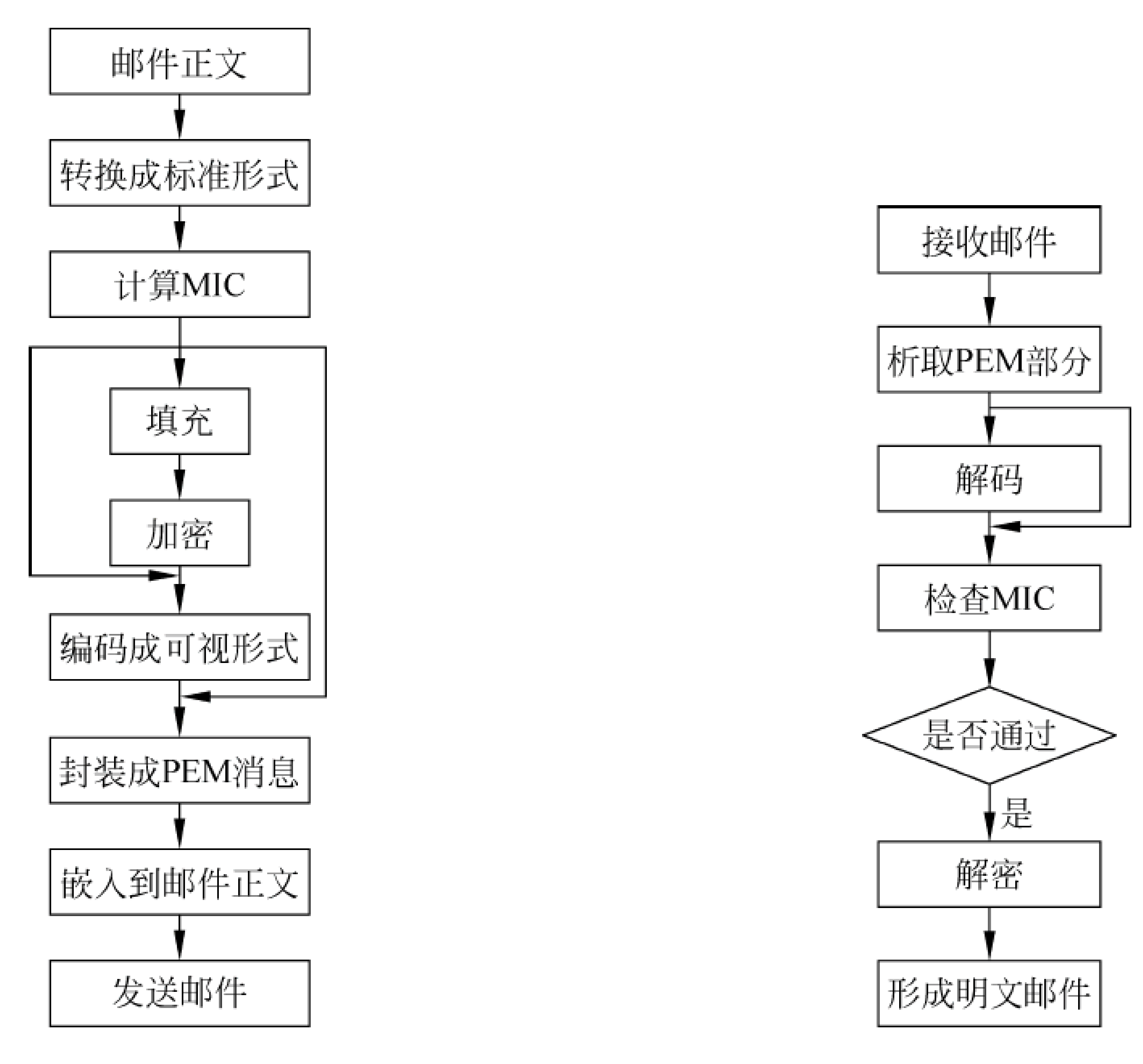


图 9-7　PEM 邮件加密签名生成过程　　**图 9-8　PEM 邮件接收和验证过程**

PEM 与 PGP 的最大区别在于 PEM 采用了基于证书的密钥管理体制。证书是数字签名、身份认证和密钥管理等各种保密措施的综合运用,它提供的安全性明显高于 PGP 等非证书密钥管理体制。虽然 PEM 标准综合运用了各种安全技术,但基于证书的密钥管理仍有不足之处。在申请证书时,用户需将自己的公钥置于申请书中形成申请证书报盘。由于证书申请不能加密,如果在申请时被第三方截获,用自己的公钥代替申请证书中的公钥,那么以后发给证书申请者的所有邮件都可能被第三者截获并解密。

9.4　PKI

PKI(Pubic Key Infrastructure)是一种遵循标准的利用公钥加密技术为电子商务的开展提供一套安全基础平台的技术和规范。用户可利用 PKI 平台提供的服务进行安全通信。使用基于公钥技术系统的用户建立安全通信信任机制的基础是:网上进行的任何需要安全

服务的通信都是建立在公钥的基础之上的,而与公钥成对的私钥只掌握在与之通信的另一方。这个信任基础是通过公钥证书的使用来实现的。公钥证书是某个用户的身份与他所持有的公钥的结合,在结合之前由一个可信任的权威机构 CA 来证实用户的身份,然后对该用户身份及对应公钥相结合的证书进行数字签名,以证明其证书的有效性。

PKI 必须具有权威认证机构 CA 在公钥加密技术基础上对证书的产生、管理、存档、发放以及作废进行管理的功能,包括实现这些功能的全部硬件、软件、人力资源、相关政策和操作程序,以及为 PKI 体系中的各成员提供全部的安全服务。例如实现通信中各实体的身份认证、保证数据的完整、抗否认性和信息保密等。

PKI 的基础技术包括加密、数字签名、数字摘要和数字信封等。

9.4.1 加密

采用密码技术对信息进行加密,使它成为密文。接收方收到密文后再对它进行解密,将密文还原成原文。

加密和解密过程需要算法和密钥。算法是加密或解密的一系列过程,在这个过程中需要一串数字,这串数字就是密钥。目前,在电子商务中普遍采用对称密钥加密系统和非对称密钥加密系统。这两种加密系统已在本文 3.3 节和 3.4 节中详细阐述过,此处不再解释。

9.4.2 数字签名

数字签名是指使用密码算法对待发的数据(报文、票证等)进行加密处理,生成一段信息附在原文上一起发送。这段信息类似于现实中的签名或印章,接收方对其进行验证来判断原文真伪。数字签名的用途是提供数据完整性保护和抗抵赖功能。该部分内容已在本文的 3.7 节中详细阐述过,此处不再解释。

9.4.3 数字信封

数字信封是信息发送端用接收端的公钥加密一个秘密密钥(SK),只有指定的接收端才能打开信封取得秘密密钥,用它来解开传输来的信息。过程如下:

(1) 要传输的信息 M 经 Hash 函数运算得到一个信息摘要 MD,MD=Hash(M)。

(2) MD 经发送方 A 的私钥 PVA 加密后得到一个数字签名 DS。

(3) 发送方 A 将信息明文 M、数字签名 DS 及其公钥 PKA 三项信息通过对称算法,以 DES 加密密钥 SK 进行加密得加密信息 E。

(4) A 在传输信息之前,必须先得到 B 的公钥 PKB,用 PKB 加密秘密密钥 SK,形成一个数字信封 DE。

(5) E+DE 就是 A 所传输的内容。

(6) 接收方 B 用自己的私钥 PVB 解开所收到的数字信封 DE,从中解出 A 使用的 SK。

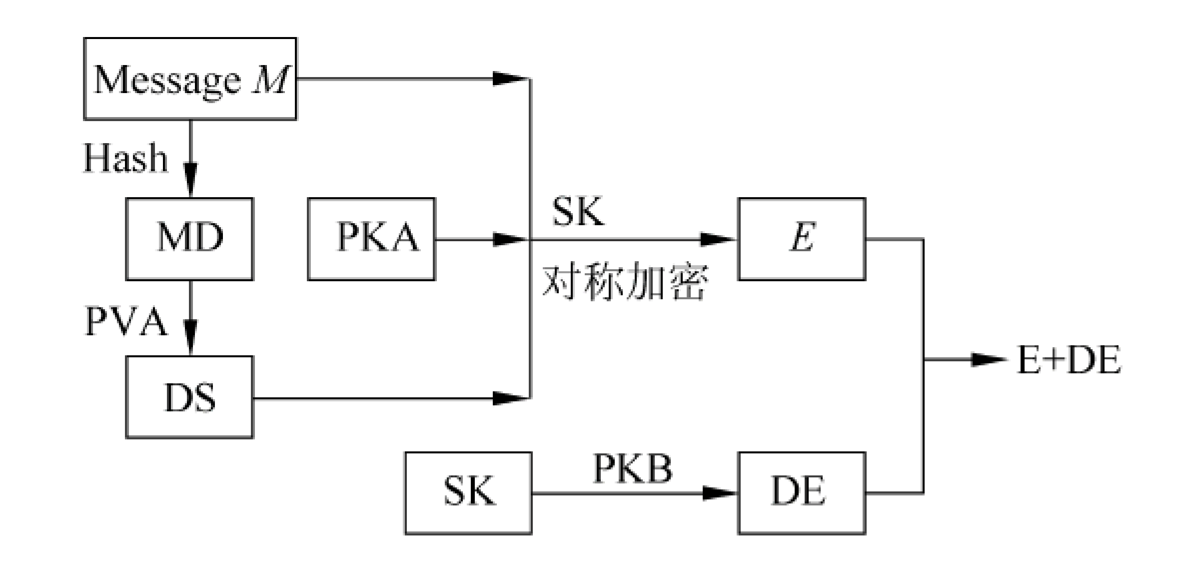
(7) B 用 SK 将 E 还原成信息明文 M、数字签名 DS 和 A 的公钥 PKA。

(8) 用 A 的公钥 PKA 将数字签名还原成信息摘要 MD。

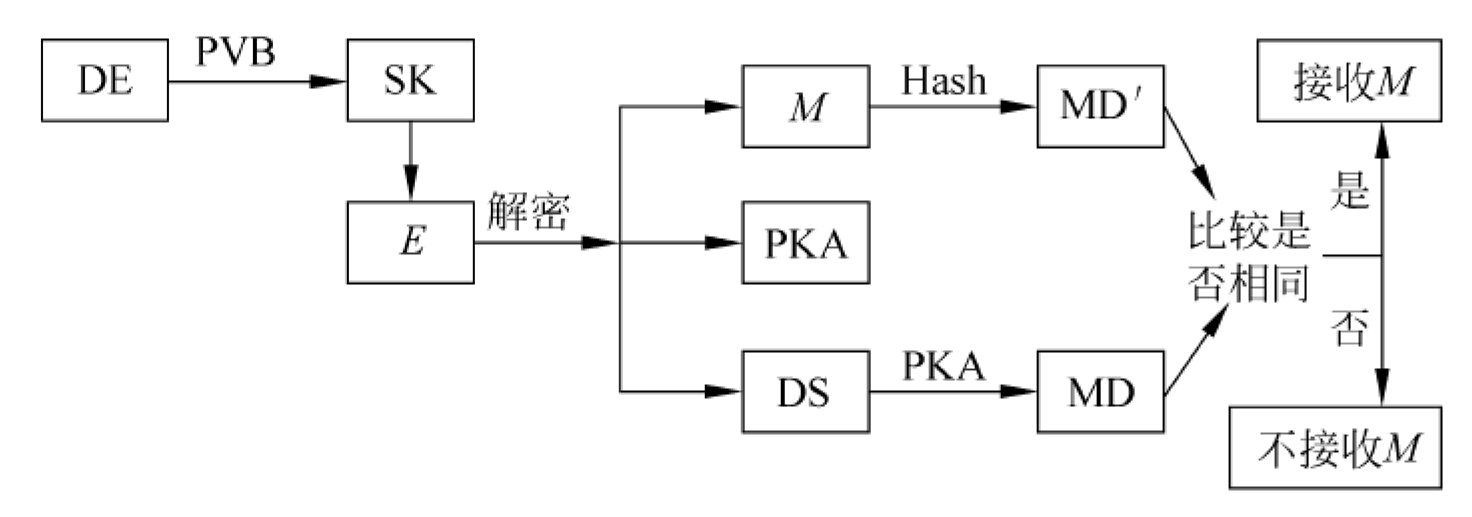
(9) B 再用收到的信息明文 M,通过 Hash 函数运算得到一个新的信息摘要 MD′。

(10) 比较已还原的 MD 和新产生的 MD′是否一致,若一致即可确认无误,否则不接收。

以上过程如图 9-9 所示。



(a) 发送方A构造数字信封



(b) 接收方B拆封数字信封

图 9-9　信息传输过程中数字信封的使用

9.4.4　数字摘要

数字摘要采用不能被解密的 Hash 函数，将原文通过特定的变换形成 128bit 的摘要。发送方将原文和摘要一同发送，接收方收到消息后用相同的 Hash 函数对收到的原文产生一个摘要，将其与收到的摘要对比。二者若相同则说明收到的信息是完整的，在传输过程中没有被修改，否则说明原文已被改动。详细内容请查阅 3.6 节。

9.5　电子邮件安全的防范措施

电子邮件的安全有两层含义：一是邮件可能给系统带来的不安全因素，二是邮件内容本身的隐私性。针对这两种情况，要有目的地增加邮件规则和系统安全方面的设置。

1. 防范欺诈邮件

欺诈性邮件一般多出现在 HTML 格式的电子邮件中。如果接收到的邮件格式是 HTML，那么它就会以 HTML 格式在用户的电子邮件客户端显示出来，这就意味着电子邮件客户端也存在着和 Web 浏览器一样受到某种 HTML 威胁的隐患。

有些欺诈性的电子邮件内容是某个网站需要用户查询一些信息，并给出一个链接，单击这个链接之后会出现可疑的登录界面。在这类邮件里，发件人通常是假冒的邮件地址。邮件内容声称该网站正在更新文件或账户，通过邮件询问用户的信息，需要收件人输入登录名与密码等信息。

2. 防范危险附件

每一份电子邮件附件都是对计算机安全性的潜在威胁。例如，W32. Gibe@mm 蠕虫是以附件的形式到达计算机的，它宣称自己是一个叫做 Q216309. exe 的 Microsoft 安全更新文件；LoveLetter 蠕虫是通过名为 LOVE-LETTER-FOR-YOU. TXT. vbs 的附件来传播的，这个文件初看上去像是人们非常熟悉的. txt 文件，但是最后的扩展名才是程序的实际关联的文件类型。因此用户应该对预料之外的附件保持警惕，无论该附件来自何处，即使看上去像是某个熟悉并可靠的联系人发送的。无论附件的来源是什么，除非它通过了最新的反病毒程序检查，否则不要在电子邮件程序中运行或打开该附件。应该把自己反病毒程序设置成"当文件被创建或保存时扫描该文件"，也可以在打开附件之前手动地用反病毒程序对附件进行扫描。

3. 安全使用基于 Web 的电子邮件

hotmail、yahoo、163 等这些基于 Web 的电子邮件系统为广大用户提供了一个方便发送和接收电子邮件的途径。许多基于 Web 的电子邮件系统都提供了一个"记住"用户名和密码的功能。如果在公用计算机上错误地选择了简易登录选项，那么其他人都会很容易地访问到用户的密码和账户。所以要注意两点：确保系统不会把用户的登录证书保存在缓存中；不使用电子邮件系统，要确保退出登录。

4. 加强电子邮件的保密性

如同明信片在邮递的过程中无保密性一样，用户发送和接收的每一封电子邮件也都没有什么保密性可言。Internet 上的有些人可以获得电子邮件的详细内容，不管它是文字、图片、音乐，还是其他任何类型的文件。在发送邮件时，电子邮件并不是直接发送到了对方的电子邮件信箱里，而是会经过数量不可预知的中间服务器。任何人只要能访问到该路径上的任何服务器，就都可以读到正在传输的信息内容。和纸质邮件一样，电子邮件的传输也与距离有关，两个电子邮件信箱之间的中间服务器节点越少，被人偷看的可能性就越低，这里的节点就相当于纸质信件的转发邮局。得到电子邮件信息虽然不能像翻开明信片那样简单，但也并不会多难。因此对于需要保密的邮件，采用数字证书帮助安全发送是最常见的方法。数字证书又名数字标识(Digital ID)，它提供了一种在 Internet 上进行身份验证的方法，用来标识和证明网络通信双方身份的数字信息文件。数字标识由公钥、私钥和数字签名 3 部分组成。在邮件中添加数字签名时，就会把数字签名和公钥加入到邮件中。因此在发送加密邮件之前，用户的通信簿必须包含对方的数字标识，才能使用他们的公用密钥来加密邮件。当收件人收到加密邮件后，用自己的私钥解密邮件之后才能阅读。还可以利用软件来对邮件进行加密。HotCrypt 是一个加密电子邮件信息的程序，也可以加/解密文本文件。HotCrypt 采取了先进的加密算法，可以有效地保障数据安全，它支持任何邮件程序或其他文件编辑窗口，通过热键即可快速加密，方便易用。

5. 邮箱炸弹的防范

邮件炸弹的原理是向有限容量的信箱投入足够多或足够大的邮件使邮箱崩溃。这类邮件炸弹很多，例如 Nimingxin、Quickfyre、Amail、Emailbomb、Upyours 系列和雪崩等，它们都能向用户信箱连续发送匿名邮件。炸弹邮件的使用也很简单，与平时书写邮件相同，填上收信人的 E-mail 地址、输入要发送的次数、选择 SMTP 主机、随意填上自己的地址，按"发

信"就开始发送炸弹了。

用户可以使用如下方法来尽可能地避免邮件炸弹的袭击：

(1) 不随意公开自己的信箱地址。

(2) 隐藏自己的电子邮件地址。例如将 aahy@sohu.com 在输入时改成 aahy.sohu.com 来避免一些邮箱自动搜索软件的识别。

(3) 谨慎使用自动回信功能。自动回信功能设计的初衷很好，但也有可能被利用来制造邮件炸弹。一旦发送方发来一封信而收件方没有及时收取，邮件服务系统就会按收件方事先的设定自动给发信人回复一封确认信。如果发信人也使用自动回信功能，就会又发给对方一封确认信，于是这种自动回复的邮件就会在两者之间重复发送，直到使双方的邮箱崩溃为止。

思　考　题

(1) 什么是电子邮件安全？什么是信息安全？

(2) 电子邮件安全与信息安全有什么联系？

(3) 怎样对信息进行保密？

(4) 加密算法的分类？这两者有何区别？

(5) 当前电子邮件面临的安全问题有哪些？

(6) 提高电子邮件安全的机制有哪些？

(7) 什么是 PGP 技术？PGP 有哪些业务及各自之间的区别？什么是数字签名？

(8) PGP 如何对消息进行处理？

(9) 会话密钥有哪些分类？如果接收方存在多个密钥对，如何恢复会话密钥及确定所用的加密算法？

(10) 在传输和接收信息时，密钥环有何作用？

(11) PGP 的安全功能和 S/MIME 的安全功能有何异同？

(12) PKI 有哪些技术？解密是否是加密的逆过程？

(13) 电子邮件安全的防范措施有哪些？

(14) 病毒、蠕虫和特洛伊木马从一个计算机传染到另一个计算机时，电子邮件的附件起了什么作用？如何进行防范？

参 考 文 献

[1] 江增世.安全邮件系统的研究与实现.电子科技大学硕士学位论文，2001.

[2] 王育民，刘建生.通信网的安全——理论与技术.西安：西安电子科技大学出版社，1999.

[3] RFC822：SMTP PROTOCOL.SPECIFICATION，http://www.ietf.org/rfc/rfc0822.txt.

[4] 卢开澄.计算机密码学——计算机网络中的数据保密与安全.北京：清华大学出版社，1998.

[5] 杨义先，等.网络信息安全与保密.北京：北京邮电大学出版社，1999.

[6] 任帅，沈思. PGP 在电子邮件安全中的应用研究.河南教育学院学报，2010，(3)：30～31.

[7] 董勇智. 电子邮件的保护神——PGP安全准则. 大众科学,2007,(2):86~87.
[8] William Stallings. 密码编码学与网络安全. 北京:电子工业出版社,2007.
[9] BehrouzA Forouzan. 密码学与网络安全. 北京:清华大学出版社,2009.
[10] 金鑫. PGP加密技术的安全性分析及改进. 中国石油大学硕士学位论文,2004.
[11] 魏洪波,周建国,梁毅. 几种电子邮件安全协议安全性分析. 网络与应用,2005(5):24~25.
[12] 胡雪梅,罗杰红. 提高电子邮件安全性的方法. 计算机时代,2002,(7).

第 10 章　无线网络安全

本章学习目标

随着无线局域网技术及其应用的迅速发展，无线网络通信的安全问题成为制约其发展的一个主要因素。本章简单介绍无线网络的基本概念及无线局域网的组织结构，详细介绍当前无线网络所面临的安全威胁及其防范措施，着重介绍无线网络的代表性技术 IEEE 802.11 安全和蓝牙安全。

通过对本章的学习，应掌握以下内容：

(1) 无线网络的基本技术概述。

(2) 无线网络的安全技术。

(3) 无线网络安全隐患和防范措施。

(4) 无线网络 IEEE 802.11 和蓝牙的安全。

近年来，伴随着 Internet 蓬勃发展的步伐，另一种联网方式已茁壮成长，这就是无线网络。目前，随着无线网络应用的日益广泛，无线网络的安全问题也越来越受到人们的普遍关注。无线网络借助于无线电波传输信息，而无线电波无处不在，且具有很强的穿透力，可在空气中任意传播，因此只要在安全控制和管理上稍有不慎就会造成很大的安全隐患。

无线网络因其能够提供比 3G 网络大得多的带宽以及 IEEE 802.11b 标准的广泛采用，而成为目前无线网络通信技术的主流。本章将从无线网络的技术概述入手，介绍无线网络的基本概念和无线局域网的组成结构与设备等；重点讲述无线网络安全的基本技术，并分析无线网络的安全隐患及防范措施；最后再针对无线网络的两个重要标准 802.11 和蓝牙的安全性进行分析介绍。

10.1　无线网络安全的基本概念

10.1.1　无线网络技术概述

无线网络(Wireless LAN，WLAN)是指用无线电波作为信息传输媒介的计算机网络系统。无线网络的范围很广泛，既包括远距离无线连接的全球语音和数据业务网络，也包括为近距离无线连接进行优化的红外线技术以及射频技术。用户可以使用计算机透过区域空间的无线网卡(Wireless Card)结合访问接入点(Access Point)进行区域无线网络连接。

无线网络是有线网络的延伸与扩展，两者最大的差别就是传输介质的不同。无线网络借助空中的无线电波，而有线网络则采用铜线或光纤等介质传输信号。无线网络与有线网络相比具有明显的优势，主要表现在：

(1) 无线局域网无须受限于网络可连线端点数的多少，可以轻松方便地在无线局域网中增加新的使用者数目，可扩展性强。

(2) 使用无线局域网时不受限于网络线的长短与插槽数目,可以节省有线网络布线所需的人力与物力成本,因此在网络结构开销上也具有一定的优势。

(3) 除了可以免去实体网络线布线的困扰之外,在网络发生错误时不用慢慢寻找损坏的线路,只要检查发送端与接收端的信号是否正常即可。

(4) 相比较于时下流行的 GPRS 手机与 CDMA 手机,无线局域网具有高速宽频上网的特性,它可提供 11 Mbps 的速度,大约是一般调制解调器(Modem 56Kbps)传输速度的 200 倍,可满足用户对大量的图像、影音传输业务的需求。

(5) 无线网络所使用的频段基本上位于 ISM 2.4GHz 的高频范围,与日常生活中所使用的电器设备之间不会产生相互干扰,而且无线网络本身共有 12 个信道可供调整,因此不会出现自然干扰的现象。

虽然无线局域网在使用的灵活性、便利性上具有非常大的优势,也能节省一定的网络结构开销,但与传统有线网络比较会发现无线局域网也存在一定的安全问题,主要体现在如下几个方面:

(1) 无线网络的信道是开放的,因此不可避免地会出现攻击者的非法接入、窃听、内容篡改以及内容转发。

(2) 局域网使用电磁波作为传输媒介,电磁波能够穿过天花板、玻璃、墙壁等物体,因此在其所服务的区域中,任何无线客户端都可以接收到信息,当然也包括那些并不希望接收数据的客户端。

(3) 无线电波在空气中的传播会因为地形、天气等多种原因造成信号衰减等问题,进而导致信号丢失,所以无线网络的信号稳定性还有待加强。

10.1.2 无线网络分类

无线网络根据数据传输范围的不同,可以分为无线个域网、无线局域网、无线城域网和无线广域网。

1. 无线个域网

无线个域网(Wireless Personal Area Network,WPAN)是一种采用无线连接的个人局域网,它应用于电话、计算机、附属设备以及小范围内的数字设备之间的通信,其工作范围一般在 10m 以内。支持个人局域网的无线技术包括蓝牙(Bluetooth)、IrDA、HomeRF 和 ZigBee 等,其中蓝牙技术的应用最为广泛。WPAN 位于整个网络链的末端,用于实现同一地点终端与终端间的连接。

2. 无线局域网

无线局域网(Wireless Local Area Network,WLAN)是个域网的延伸,它是基于以太网技术在本地(例如公司、校园、机场等场所)创建的无线连接。无线局域网络是相当便利的数据传输系统,它利用射频(Radio Frequency,RF)技术,取代旧式的有线局域网,解决了在铺设缆线困难的场所实现局域网络的连接问题,以及在现有 LAN 设施的基础上实现不受时空限制的网络连接功能。

3. 无线城域网

无线城域网(Wireless Metropolitan Area Network,WMAN)是连接数个无线局域网的

无线网络形式。无线城域网主要用于解决城市区域内网络的接入问题，覆盖范围为几千米到几十千米，除了提供固定的无线接入外，还能提供具有移动性的接入能力。由于 WLAN 的总体设计及其提供的特点并不能很好地适用于室外应用，当其用于室外时，在带宽和用户数等方面将受到限制，同时还存在着通信距离等其他一些问题。

4. 无线广域网

无线广域网（Wireless Wide Area Network，WWAN）是指通过远程公共网络或专用网络建立的无线连接的网络。这些连接可以通过一些天线基站或卫星系统覆盖范围广大的地理区域，例如不同的国家、城市和地区。与无线个域网、无线城域网及无线局域网相比，无线广域网更加强调的是快速移动性。WWAN 技术使得笔记本计算机或者其他设备装置在蜂窝网络覆盖范围内可以在任意地方连接到 Internet。

10.1.3　无线网络协议

目前，WWAN 采用的主要技术以 2G、3G 为代表，并逐步向超 3G 过渡。2G 网络系统为全球数字移动电话系统（GSM）、网络数字包数据（CDPD）和码分多址（CDMA）技术的系统，是目前普遍采用的 WWAN 网络技术。如今该技术正在从 2G 网络系统过渡到 3G 网络系统，以解决目前 2G 系统存在的受限漫游功能和系统互不兼容的问题。3G 技术将执行全球标准并提供全球漫游功能。

3G 技术是由国际电信联盟（International Telecommunication Union，ITU）于 1985 年提出的，后来更名为 IMT-2000。它是一种能提供多种类型、高质量、多媒体业务的全球漫游移动通信网络。3G 网络的信息覆盖区域为国家级，且使用频带需要国家许可，因此现在的发展还不能像 WLAN 技术一样普遍。

无线网络的代表性技术有 IEEE 802.11 系列、蓝牙技术、HiperLAN（高性能无线局域网）、HomeRF、IrDA 和 ZigBee 等。

1. IEEE 802.11 系列

在 IEEE 802.11 系列中，目前应用最广泛的是 802.11b，又称为 Wi-Fi（Wireless Fidelity，无线保真），它使用开放的 2.4GHz 直接序列扩频（Direct Sequence Spread Spectrum，DSSS），最大数据传输速率为 11Mbps，可穿越障碍物，非直线传播时的传输最大范围为室外 300m，室内 100m。其优势在于价格低廉，可与 AP 的动态安全加密相匹配，但与 802.11a/g 相比，其速率较低。

802.11a 扩充了 802.11 标准的物理层，工作在 5GHz 无线频段，从而避开了拥挤的 2.4GHz 频段，采用正交频分复用技术（Orthogonal Frequency Division Multiplexing，OFDM），物理层传输速率可达 54Mbps，传输层可达 25Mbps，传输距离和 802.11b 差不多。其优势在于传输速率高且受干扰少，但价格相对较高。

802.11g 的出现使得 WLAN 提供低价高速的产品成为可能。该标准工作在 2.4GHz 无线频段，最高数据传输速率为 54Mbps，与 802.11a 相当，且与 802.11b 保持兼容。802.11g 与 802.11a 一样，也采用正交频分复用技术，但使用的频段与 802.11b 相同（2.4GHz），因而其传输距离更远，在室内可达到 150m。

表 10-1 列出了 3 种 IEEE 802.11 标准的比较。

表 10-1 3 种 IEEE 802.11 标准比较

	802.11a	802.11b	802.11g
工作频率	5GHz	2.4GHz	2.4GHz
传输频率	54Mbps	11Mpbs	54Mbps
调制类型	OFDM(正交频分复用)	DSSS(直接序列扩频)	OFDM(正交频分复用)

2. 蓝牙技术

蓝牙是一种支持设备短距离通信(一般 10m 内)的无线技术。它工作在 2.4GHz ISM(即工业、科学、医学)频段,目前可支持 1Mbps 的数据传输速率,支持数据和语音业务,应用于掌上计算机、笔记本计算机和移动电话等便携型移动终端设备,可简化这些设备与 Internet 之间的通信,使数据传输更加高效快速。

3. 高性能无线局域网

HiperLAN(High Performance Radio LAN)是在欧洲应用的无线局域网通信标准的一个子集。它有两种规格:HiperLAN/1 和 HiperLAN/2。这两种标准均被欧洲电信标准协会(ETSI)采用。HiperLAN 标准提供了类似于 IEEE 802.11 无线局域网协议的性能。HiperLAN/1 标准采用 5GHz 射频频率,采用 GMSK(调制前高斯滤波的最小频移键控)技术。HiperLAN/2 同样采用 5GHz 射频频率,上行速率达到 54Mpbs,采用的则是 OFDM(正交频分复用)技术,而且可以同 3G 标准兼容。

4. HomeRF

HomeRF 无线标准是由 HomeRF 工作组开发的开放性行业标准,目的是在家庭范围内实现计算机与其他电子设备之间的无线通信。它采用开放的 2.4GHz 频段,采用跳频扩频技术,跳频速率为 50 跳/s,共有 75 个宽带为 1MHz 的跳频信道。HomeRF 是对现有无线通信标准的综合和改进:当进行数据通信时,采用 IEEE 802.11 规范中的 TCP/IP 协议;当进行语音通信时,则采用数字增强型无绳通信标准。

HomeRF 的特点是安全可靠、成本低廉、不受墙壁和楼层的影响,而且无线电干扰小,能支持流媒体。但是它与 802.11b 标准不兼容,并且所占频段与 802.11b 和蓝牙一样为 2.4GHz,所以在应用范围上有一定的局限性,一般在家庭网络中使用。

5. IrDA

IrDA 是一种利用红外线进行点对点通信的技术,由红外线数据标准协会(Infrared Data Association)制定的无线协议。目前 IrDA 的通信最高速率为 4Mbps,且要求通信距离在 1m 以内,同时在点对点通信时要求接口对准角度不能超过 30°。红外信号要求视距传播,方向性强,不易对其他系统产生干扰,并且难以窃听,安全性高。

IrDA 技术作为一种无线技术,旨在让那些只需要收发少量信息的设备进行通信。因为这种技术比较便宜,所以许多个人设备都集成了这种技术,包括笔记本计算机、手持设备、计算机外设等。由于红外线受日光、环境照明等影响较大,一般要求发射功率较高。且红外应用存在着上述特定要求,所以同蓝牙技术相比,IrDA 技术存在着许多局限。

6. ZigBee

ZigBee 是基于 IEEE 802.15.4 的低功耗个域网协议。ZigBee 是一种新兴的近距离、低

复杂度、低功耗、低数据速率且低成本的无线网络技术，它是一种介于无线标记和蓝牙之间的技术提案。ZigBee 在室内通常能达到 30～50m 作用距离，在室外甚至可以达到 400m。

ZigBee 的传输频带有 3 种：868MHz，传输速率为 20Kbps，该频带适用于欧洲；915MHz，传输速率为 40Kbps，该频带适用于美国；2.4GHz，传输速率为 250Kbps，该频带全球通用。

10.1.4　无线网络设备

在无线局域网里，常见的网络设备主要有无线网卡、无线 AP、无线路由器、无线网桥、无线天线等。

1. 无线网卡

无线网卡的作用类似于以太网中的网卡，起到信号接收的作用，作为无线局域网的接口，实现与无线局域网的连接。无线网卡根据接口类型的不同主要分为 3 种：PCMCIA 无线网卡、PCI 无线网卡和 USB 无线网卡。PCMCIA 无线网卡仅适用于笔记本计算机，支持热插拔，可以非常方便地实现移动无线接入。PCI 无线网卡是在 PCI 转接卡上插入一块普通的 PCMCIA 卡，适用于普通的台式计算机。USB 无线网卡适用于笔记本和台式机，支持热插拔，如果网卡外置有无线天线，那么 USB 无线网卡就是一个比较好的选择。

2. 无线 AP

无线 AP(Wireless Access Point)是移动计算机用户进入有线网络的接入点，主要用于宽带家庭、大楼内部以及园区内部，典型距离覆盖几十米至上百米，目前主要技术为 802.11 系列。单纯无线 AP 就是一个无线的交换机，提供无线信号发射机的功能。它的工作原理是将网络信号通过双绞线传输，经过 AP 产品的编译，将电信号转换成无线电信号发送出去，形成无线网的覆盖。

AP 是传统的有线局域网与无线局域网之间的桥梁，任何一台装有无线网卡的计算机都可以通过 AP 来实现有线局域网或广域网的信息资源共享。此外，AP 本身又兼具网管的功能，可对接有无线网卡的计算机进行控制和管理。AP 主要分为不带路由功能、仅提供无线信号发射功能的普通 AP 和带路由功能、可以实现为拨号接入 Internet 的 ADSL 等提供自动拨号功能的 AP 两种。

3. 无线路由器

无线路由器是带有无线覆盖功能的路由器，它主要用于用户上网和无线覆盖。无线路由器可以看成是普通无线 AP 和宽带路由器合二为一的扩展型产品，它不仅具备普通 AP 所具有的所有功能，例如支持 DHCP(Dynamic Host Configuration Protocol，动态主机设置协议)客户端、支持 VPN、防火墙、WEP 加密等，而且还具有网络地址转换(NAT)功能，支持局域网用户的网络连接共享。这样便可实现家庭无线网络中的 Internet 的 ADSL 和小区宽带的无线共享接入。

4. 无线网桥

从作用上来理解无线网桥，它可以用于连接两个或多个独立的网络段，这些独立的网络段通常是位于不同的建筑内，相距几百米到几十千米，所以说它可以广泛应用于不同建筑物间的网络互联。同时，根据协议不同，无线网桥又可以分为 2.4GHz 频段的 802.11b 或

802.11g 以及采用 5.8GHz 频段的 802.11a 无线网桥。无线网桥有 3 种工作方式,即点对点、点对多点和中继连接,特别适用于城市中的远距离通信。

5. 无线天线

当计算机与无线 AP 或其他计算机相距较远时,随着信号的减弱,或者传输速率明显下降,或者根本无法实现与 AP 或其他计算机之间通信,此时就必须借助于无线天线对所接收或发送的信号进行增益放大。

无线天线有多种类型,不过常见的有两种,一种是室内天线,另一种是室外天线。室内天线的优点是方便灵活,缺点是增益小,传输距离短。室外天线的类型比较多,一种是锅状的定向天线,一种是棒状的全向天线。室外天线的优点是传输距离远,比较适合远距离传输。

10.1.5 无线网络的应用模式

无线局域网具有 Ad-hoc 和 Infrastructure 两种网络应用模式。

1. Ad-hoc 模式

Ad-hoc(自组织结构)是一种特殊的无线网络应用模式,如图 10-1 所示。一组计算机如果不需要访问网络资源,可以接上无线网卡相互连接,即可建立临时网络,共享资源,而无须通过 AP。在这种模式下,客户端与客户端之间可以在网络内无须通过 AP 而直接通信。Ad-hoc 模式和有线网络的直连双绞线概念类似,是 P2P 的连接,因此也就无法与其他网络沟通了。一般无线终端设备像 PMP、PSP、DMA 等应用的就是 Ad-hoc 模式。

Ad-hoc 的原理是网络中的一台计算机主机建立点对点连接相当于虚拟 AP,而其他计算机就可以直接通过这个点对点连接进行互联与共享。由于省去了 AP,Ad-hoc 无线局域网的架设过程十分简单,但它的有效传输距离只有 40m 左右,因此这种模式只适合一些简单的或者临时性的网络互联需求。



图 10-1 Ad-hoc 网络模式

2. Infrastructure 模式

Infrastructure(基础结构)模式为一种整合有线与无线局域网结构的应用模式,通过此模式,可达成有线和无线网络之间的资源共享,如图 10-2 所示。与 Ad-hoc 不同的是配备无线网卡的计算机必须通过 AP 来进行无线通信。客户端与 AP 之间在传输数据之前必须建立关联,建立关联时必须拥有相关的授权才能实现信息交换。此时 AP 以特定的频率发送标识管理帧,客户端收到后,通过发送授权帧建立授权。授权成功后,客户端又发送关联帧,最后 AP 回应关联。

Infrastructure 模式可以分为无线 AP 无线网卡模式和无线路由器无线网卡模式。无线 AP 无线网卡模式中当网络中存在着一个 AP 时,无线网卡的覆盖范围将变为原来的两倍,而且可以增加无线局域网所容纳的网络设备。但是无线 AP 的作用类似于有线网络的集线器,只有单纯的无线覆盖功能。无线路由器无线网卡模式是现在很多家庭都采用的无线组网模式,此时的无线路由器就相当于一个无线 AP 加上路由的功能。



图 10-2　Infrastructure 网络模式

10.2　无线网络安全技术

安装无线局域网如同在任何地方都放置以太网端口一样,任何人都可以访问网络。在一个服务集内,目前还没有办法直接将传输数据指向预定接收端。因此与有线线缆的组网方式不同,在无线局域网中,只要某台设备工作在与无线局域网相同的频带上,那么它就有可能在满足某些特定条件的情况下对传输信息进行窃听或者对传递信息的信号实施蓄意干扰,这不仅影响无线网络自身的安全,还危及与之相连接的有线网络安全。为了阻止未经授权用户对无线网络的非法访问,并阻止对无线局域网数据流的非法侦听,无线网络采用了一些有效的安全技术。

1. 无线跳频扩频技术

扩展频谱技术又称扩频技术,是近年发展非常迅速的一种技术,将其用于无线局域网中能使系统的各项性能得到改善,它已成为无线局域网中不可缺少的一种技术。扩展频谱技术在 50 年前第一次被美国军方公开,用来进行保密传输。扩展频谱发送器用一个非常弱的功率将信号在一个很宽的频率范围内发射出去。扩展频谱的实现方式有多种,最常用的两种是直接序列扩频(Direct Sequence Spread Spectrum, DSSS/DS)和跳频扩频技术(Frequency Hopping Spread Spectrum,FHSS/FH)。

1) 直接序列扩频

直接序列扩频(DS)就是用高速的伪随机序列(PN 码)与信息码序列模 2 加后生成的复合码序列去控制载波,从而获得直接序列扩频信号,如图 10-3 所示。

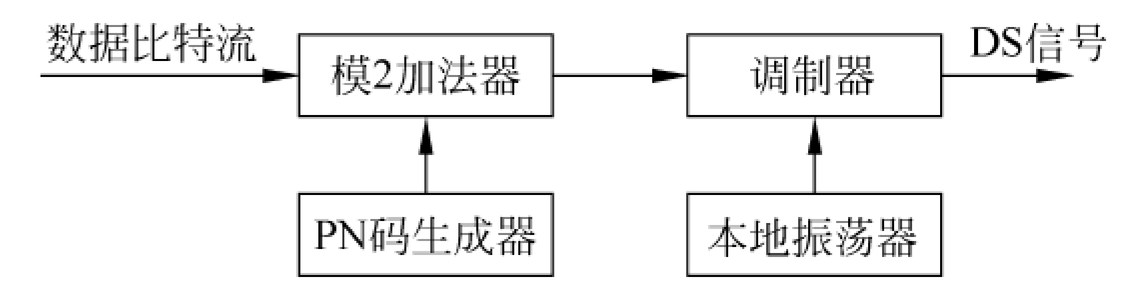


图 10-3　简单的直接序列扩频原理框图

2) 跳频扩频

跳频扩频(FH)是使发射频率在一组预先指定的频率上按照编码序列所规定的顺序离

散地跳变,从而扩展发射频谱。跳频扩频一般用伪随机序列控制频率合成器后构成跳频指令,根据跳频指令随机选择发送频率。在接收端使用与发送端相同的伪随机序列发生器构成跳频指令去控制频率合成器,从而恢复出原信号,如图 10-4 所示。

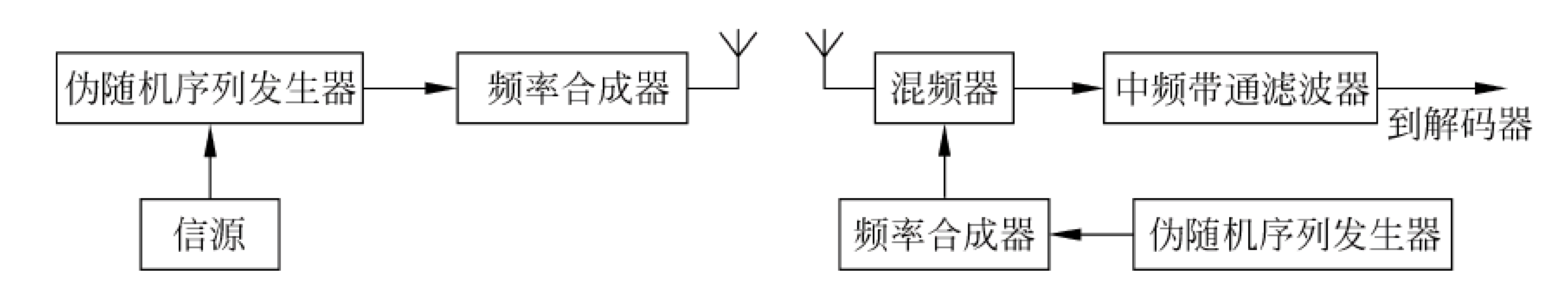


图 10-4 跳频扩频系统原理框图

在跳频方式扩展信号中,ISM 频段中的 2.400～2.483GHz 被分成 79 个不同的频道。传输采用一组随机序列的频点来发送。如果不知道在每一个频道上的停留时间和它的跳频图案,对于一个企图非法加入的工作站来说是不可能接收和释译数据的。使用不同的跳频图案、停留时间和频道数可以允许两个距离很近的局域网同时存在,并且没有相互干扰和窃听数据的可能。跳频扩频技术也是无线网络中抗干扰的主要方法。

2. 扩展服务集标识号

每个无线设备都内置了一个 32 位的扩展服务集标识号(ESSID),对于任何一个可能存取接入点的适配器,接入点首先要决定这个适配器是否属于该网络。只有当 AP 和无线终端的 ESSID 相匹配时,AP 才接受无线终端的访问并提供网络服务;如果不符合则拒绝提供服务。如果需要在一个网络上有不同的网段,那么可以编写不同的 ESSID。如果需要支持移动用户和扩大带宽而连接多个接入点,那么它们的 ESSID 必须设置成一致,而跳频序列应该不一样。所有这些设置都受接入点安装者密码的控制。想要推断确切的 ESSID 和跳频序列进行窃听是非常困难的。

3. 有线对等协议

无线网络安全的另一个重要方面是数据加密,一般可以通过有线对等协议(Wired Equivalent Protection,WEP)来实现。WEP 是由 IEEE 802.11 标准定义的,是最基本的无线安全加密措施,它主要应用于无线局域网中链路层数据。WEP 的主要用途是为无线局域网提供与有线网络相同级别的安全保护,保证无线通信信号的安全性、保密性以及完整性,具体表现为:

(1) 提供接入控制,防止未授权用户访问网络。

(2) WEP 加密算法对数据进行加密,防止数据被攻击者窃听。

(3) 防止数据被攻击者中途恶意篡改或伪造。

WEP 的数据帧格式如图 10-5 所示。

由图 10-5 可见,WEP 数据帧分为 3 项:32b 的 IV(初始向量)、DATA(传输数据)及 32b 的 ICV(即 32 位循环校验码)。值得注意的是,IV 是以明文方式传输的,而数据及 ICV 是以密文方式(Encrypted)传输的。

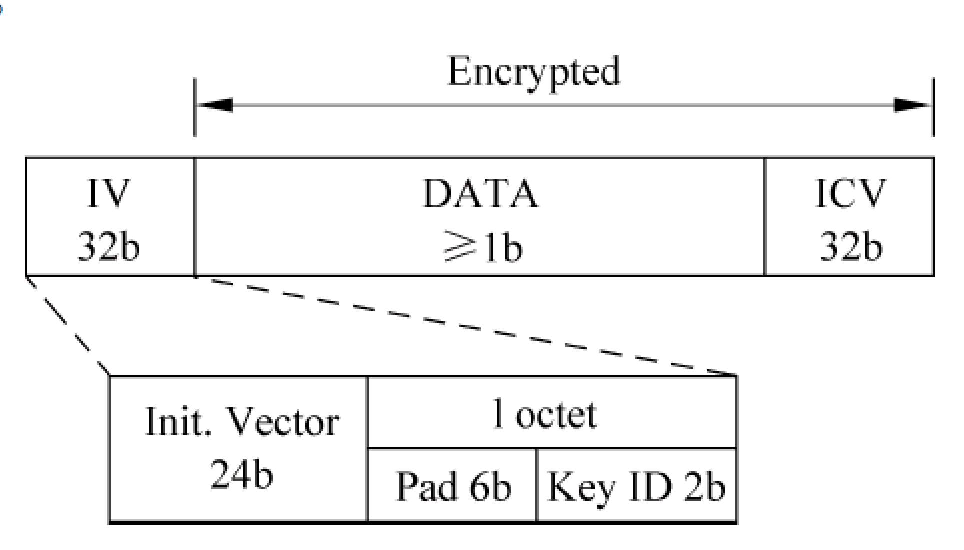


图 10-5 WEP 的数据帧格式

IV 包含 3 个子数据:24b 的初始向量值

Init. Vector、2b 的 Key ID 和 6b 的填充数据（全部用 0 填充）。其中 Init. Vector 用来构成 WEP Seed，Key ID 用于选择加密该数据帧所使用的密钥（因为 WEP 中共有 4 个密钥，所以用 Key ID 来帮助选择）。

WEP 用来阻止对无线网络的非法访问。在外传数据已加密并打包的情况下，WEP 默认停止使用。WEP 使用 BSS 上的共享密钥 RC4 加密算法，密钥长度从最初的 40b（5 个字符）变为后来的 128b（13 个字符），有些设备还可以支持 152b 或 256b 加密。WEP 加密采用静态密钥，各 WLAN 终端使用相同的密钥访问无线网络，数据在无线发射之前进行复杂的编码处理，接收端收到之后通过反向处理获取数据原值。这种加密方式能够确保数据在泄露的情况下不会暴露数据的原值。除此之外，WEP 也可实现认证功能。在加密功能启用的情况下，客户端要尝试连接 AP 时，AP 会发出一个检测挑战值包（Challenge Packet）给客户端，客户端再利用共享密钥将此值加密后送回接入点以进行认证比对，如果无误才准许存取网络的资源。

WEP 采用对称加密机理，数据的加密和解密采用相同的密钥和加密算法。WEP 加密过程依赖通信双方共享的密钥来保护传输的加密帧数据。解密过程是加密过程的简单取反。WEP 的加密算法示意图如图 10-6 所示。

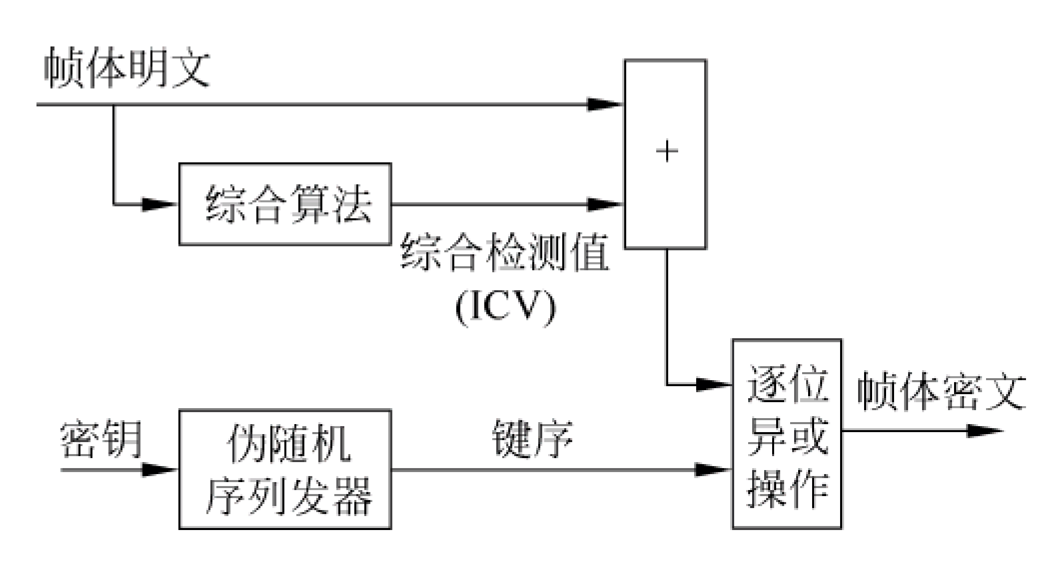


图 10-6　WEP 的加密算法示意图

4. WPA 规范

WPA（Wi-Fi Protected Access）规范是 Wi-Fi 联盟于 2002 年 11 月发布的一项代替 WEP 的过渡性的无线网络安全标准，是 IEEE 802.11i 的一个子集。它在 IEEE 802.11i 标准最终确定前，将为 WLAN 提供更强大的安全性能。WPA 的核心是 IEEE 802.1x 和临时密钥完整性协议（Temporal Key Protocol，TKIP）。

WPA 对于不同的用户和不同的应用安全需要应用不同的安全模式。例如，企业用户需要很高的安全保护，否则可能会泄露非常重要的商业机密；而家庭用户往往只是使用网络来浏览网页、收发 E-mail、打印和共享文件等，这些用户对安全的要求相对较低。为满足不同用户需要，WPA 中规定了两种应用模式。

（1）企业模式：通过使用认证服务器和复杂的安全认证机制来保护无线网络通信安全。

（2）家庭模式（包括小型办公室）：在 AP（或者无线路由器）及连接无线网络的无线终端上输入共享密钥来保护无线链路的通信安全。

5. TKIP

临时密钥完整性协议（Temporal Key Integrity Protocol，TKIP）是新一代的数据加密技

术，它与 WEP 一样是基于 RC4 加密算法的，但对现有 WEP 进行了改进，采用 128b 的密钥长度，并采用了动态会话密钥。采用 TKIP 后，IV 会以加密形式传输，黑客想要在无线网络中窃取 IV 就没那么容易了。TKIP 引入了 4 种新算法，极大地提高了加密安全强度。这 4 种新算法是：

(1) 具有序列功能的 48b 初始化向量(IV)和 IV 序列规则(IV Sequencing Rule)。

(2) 每包密钥构建(Per-Packet Key Construction)，用于实现密钥细分，即每发一个包就重新生成一个新的密钥。

(3) Michael 消息完整性代码(Message Integrity Code，MIC)。

(4) 密钥重获与分发。

6. AES

IEEE 802.11i 中还定义了一种基于高级加密标准(Advanced Encryption Standard，AES)的全新加密算法，又称 Rijndael 加密法，以实现更强大的加密和信息完整性检查。AES 是一种对称的块加密技术，密钥长度最少支持为 128b、192b、256b，分组长度 128b，算法应用于各种硬件和软件实现提供比 WEP4/TKIP 中 RC4 算法更高的加密性能。表 10-2 列出了不同密钥长度下破解 AES 加密所需进行的猜测次数。

AES 算法的一个不足之处在于它的开销远大于 RC4 算法。这是由于 AES 算法在加密和解密过程中需要进行额外的处理，与相对简单的 RC4 算法相比，这些处理要复杂得多。尽管如此，由于其在加密性能上的优势，AES 算法注定会成为无线通信加密的主流方法。

表 10-2　AES 不同密钥长度的解密次数

密钥长度	解密次数
128b 密钥	3.4×10^{38} 次
192b 密钥	6.2×10^{57} 次
256b 密钥	1.1×10^{77} 次

尽管 WLAN 采用了多种安全技术以保证网络的安全性，但因为这些安全技术本身的漏洞及无线网络安全配置的不合理性，致使无线网络仍然存在着安全问题。在下一节中将重点讨论无线网络的安全隐患及其对策。

10.3　无线网络安全问题

10.3.1　无线网络安全性的影响因素

无线局域网的安全隐患主要集中在如下几个方面：

1. 无线通信覆盖范围问题

由于无线网络设计的基础是利用无线电波来实施传输，没有明确的覆盖范围，这样就会使网络攻击者对无线电波覆盖范围内的数据流进行侦听，如果无线用户没有对传输的信息实施加密的话，那么网络攻击者就可以非常容易地窃取所有通信信息。

另外，无线网络只要在无线电波覆盖范围内就可以使用，因此对其管理和控制就没有传统有线网络那么容易。并且大多数无线局域网所使用的都是 ISM 频段，在该频段范围内工作的设备非常多，因此存在同信道和临信道以及其他设备相互之间的干扰问题。

2. 无线设备管理问题

对无线网络设备而言,在其出厂时会有一些预先设定的设定值,许多无线用户并没有对购买到的无线设备实施有效配置,这样网络攻击者就可以利用这些潜在的安全漏洞对网络实施攻击。

3. 密钥管理问题

在无线网络中并没有针对无线网络加密密钥的管理与分配机制,这样在无线网络中就会存在对密钥管理与分配的很大困难。

4. 现有 WEP 协议安全漏洞

安全领域中的一个重要规则就是没有安全措施比拥有虚假安全措施更可怕。虽然 WEP 并不能算作虚假安全措施,但是在其设计过程中确实存在许多安全漏洞。

1) 缺少密钥管理

用户的加密密钥必须与 AP 的密钥相同,并且一个服务区内的所有用户都共享同一把密钥。WEP 标准中并没有规定共享密钥的管理方案,通常是手工进行配置与维护。由于同时更换密钥的费时费力,所以密钥通常长时间使用而很少更换,倘若一个用户丢失密钥,则将殃及整个网络。

2) ICV 算法不合适

ICV 是一种基于 CRC-32 的用于检测传输噪音和普通错误的算法。CRC-32 是信息的线性函数,这意味着攻击者可以篡改加密信息,并且很容易修改 ICV,使信息表面上看起来是可信的。能够篡改加密数据包使各种各样的非常简单的攻击成为可能。

3) RC4 算法存在弱点

在 RC4 算法中,人们发现了弱密钥。所谓弱密钥就是密钥与输出之间存在超出一个好密钥所应具有的相关性。在 24b 的 IV 值中有 9000 多个弱密钥。攻击者收集到足够的使用弱密钥的包后,就可以对它们进行分析,只需尝试很少的密钥就可以接入到网络中。

5. 缺少交互认证

无线局域网设计的另一个缺陷就是状态机中用户和 AP 之间的异步性。根据标准,仅当认证成功后认证端口才会处于受控状态。但对于用户端来说并不是这样的,其端口实际上总是处于认证成功后的受控状态。而认证只是 AP 对用户端的单向认证,攻击者可以处于用户和 AP 之间,对用户来说攻击者充当成 AP,而对于 AP 来讲攻击者则充当用户端。IEEE 802.1x 规定认证状态机只接收用户的 PPP 扩展认证协议(Extensible Authentication Protocol,EAP)响应,并且只向用户发送 EAP 请求信息。类似地,用户请求机不发送任何 EAP 请求信息,状态机只能进行单向认证。从这个设计中反映出来一个信任假设,即 AP 是受信的实体,这种假设是错误的。如果高层协议也只进行单向认证的话,则整个框架就是不安全的。

10.3.2 无线网络常见攻击

目前,由于大多数的 WLAN 默认设置为 WAP 不起作用,攻击者可以通过扫描找到那些允许任何人接入的开放式 AP 来得到免费的 Internet 使用权限,并能以此发动其他攻击。无线网络中常见的攻击形式如下:

1. MAC 地址嗅探(MAC Sniffing)

检测 WLAN 非常容易,目前有一些工具可运行在 Windows 系统上或 GPS 接收器上来定位 WLAN,例如 NetStumbler、Kjsmet 可识别 WLAN 的 SSID 并判断其是否使用了 WEP,还可以识别 AP 和 MAC 地址。

2. AP 欺骗(Access Point Spoofing)和非授权访问

无线网卡允许通过软件更换 MAC 地址,攻击者嗅探到 MAC 地址后,通过对网卡的编程将其伪装成有效的 MAC 地址,进入并享有网络。

MAC 地址欺骗是很容易实现的,使用捕获包软件,攻击者能获得一个有效的 MAC 地址包。如果无线网卡防火墙允许改变 MAC 地址,并且攻击者拥有无线设备且在无线网络附近的话,攻击者就能进行欺骗攻击。欺骗攻击时,攻击者必须设置一个 AP,它处于目标无线网络附近或者在一个可被受攻击者信任的地点。如果假的 AP 信号强于真的 AP 信号,受攻击者的计算机将会连接到假的 AP 中。一旦受攻击者建立连接,攻击者就能偷窃他的密码,享有他的权限等。

因为 TCP/IP 协议的设计原因,几乎无法防止 MAC/IP 地址欺骗。只有通过静态定义 MAC 地址表才能防止这种类型的攻击。但是因为巨大的管理负担,这种方案很少被采用。只有通过智能事件记录和监控日志才可以对付已经出现过的欺骗。当试图连接到网络上的时候,简单地通过让另外一个节点重新向 AP 提交身份验证请求就可以很容易地欺骗无线网身份验证。许多无线设备提供商允许终端用户通过使用设备附带的配置工具,重新定义网卡的 MAC 地址。使用外部双因子身份验证,例如 RADIUS 或 SecurID,可以防止非授权用户访问无线网及其连接的资源,并且在实现的时候应该对需要经过强验证才能访问资源的访问进行严格限制。

3. 窃听、截取

窃听是指偷听流经网络的计算机通信的电子形式,它是以被动和无法觉察的方式入侵检测设备的。无线网络最大的安全隐患在于入侵者可以访问某机构的内部网络,即使网络不对外广播网络信息,只要能够发现任何明文信息,攻击者仍然可以使用一些网络工具,例如 Ethreal 和 TCP Dump 来窃听和分析通信量,从而识别出可以破坏的信息。使用虚拟专用网(VPN)、SSL(Secure Sockets Lave,安全套接层)和 SSH(Secure Shel1)有助于防止无线拦截。

4. 网络接管与篡改

同样因为 TCP/IP 协议设计的原因,某些技术可供攻击者接管与其他资源建立的网络连接。如果攻击者接管了某个 AP,那么所有来自无线网的通信量都会传到攻击者的机器上,包括其他用户试图访问合法网络主机时需要使用的密码和其他信息。接管 AP 可以让攻击者从有线网或无线网进行远程访问,而且这种攻击通常不会引起用户的重视,用户通常是在毫无防范的情况下输入自己的身份验证信息,甚至在接到许多 SSL 错误或其他密钥错误的通知之后,仍像是看待自己机器上的错误一样看待它们,这让攻击者可以继续接管连接,而不必担心被别人发现。

5. 拒绝服务攻击(DoS)

无线信号的传输特性和专门使用扩频技术,使得无线网络特别容易受到拒绝服务

(Denial of Service,DoS)攻击的威胁。拒绝服务是指攻击者恶意占用主机或网络几乎所有的资源,使得合法用户无法获得这些资源。这类攻击最简单的实现办法是通过让不同的设备使用相同的频率,从而造成无线频谱内出现冲突。另一个可能的攻击手段是发送大量非法(或合法)的身份验证请求。第三种手段:如果攻击者接管 AP,并且不把通信量传输到恰当的目的地,那么所有的网络用户都将无法使用网络。无线攻击者可以利用高性能的方向性天线,从很远的地方攻击无线网。已经获得有线网访问权的攻击者,可以通过发送多达无线 AP 无法处理的通信量来攻击它。此外,为了获得与用户的网络配置发生冲突的网络,只要利用 NetStumbler 就可以做到。

6. 主动攻击

主动攻击比窃听更具危害性。入侵者将穿过某机构的网络安全边界,而大部分安全防范措施(防火墙、入侵检测系统等)都安排在安全边界之外,界线内部的安全性相对薄弱。入侵者除了窃取机密信息外,还可利用内部网络攻击其他计算机系统。

7. WEP 攻击

WEP 最初的设计目的就是为了提供以太网所需要的安全保护,但其自身存在着一些致命的漏洞。在无线环境中,不使用保密措施是具有很大风险的,但 WEP 协议只是 IEEE 802.11 设备实现的一个可选项。WEP 中的 IV 由于位数太短和初始化复位设计,容易出现重用现象,从而被他人破解密钥。而对用于进行流加密的 RC4 算法,在开始 256B 数据中的密钥存在弱点,目前还没有任何一种实现方案修正了这个缺陷。此外用于对明文进行完整性校验的 CRC(Cyclic Redundancy Check,循环冗余校验)只能确保数据正确传输,并不能保证其未被修改,因而并不是安全的校验码。

IEEE 802.11 标准指出,WEP 使用的密钥需要接受一个外部密钥管理系统的控制。通过外部控制,可以减少 IV 的冲突数量,使得无线网络难以攻破。但问题在于这个过程形式非常复杂,并且需要手工操作。因而很多网络的部署者更倾向于使用默认的 WEP 密钥,这使黑客为破解密钥所做的工作量大大减少了。另一些高级的解决方案需要使用额外资源,例如 RADIUS 和 Cisco 的 LEAP,其花费是很昂贵的。

10.3.3　无线网络安全技术措施

1. 使用防火墙

从前文可知,防火墙(Firewall)是由软件、硬件构成的系统,用来在两个网络之间实施接入控制。防火墙是指隔离在本地网络与外界网络之间的一道执行控制策略的防御系统,它对网络之间传输的数据包依照一定的安全策略进行检查,以决定通信是否被允许,对外屏蔽内部网络的信息、结构和运行状况,并提供单一的安全和审计的安装控制点,从而达到保护内部网络信息不被外部非授权用户访问和过滤不良信息的目的。

防火墙的功能有两个:一个是阻止,另一个是允许。大多数情况下,防火墙的主要任务是阻止。目前防火墙技术主要包括包过滤技术、应用代理技术和状态检测技术,可以实现对输入进行筛选、防止内部信息的外泄、限制内部用户活动和对网络使用情况进行记录、监控等。

2. 修改用户名和密码

一般的无线网络都是通过无线路由器或中继器来访问外部网络的。通常这些路由器或

中继器设备制造商为了便于用户设置这些设备建立起无线网络，都提供了一个管理页面工具。这个页面工具可以用来设置该设备的网络地址以及账户等信息，为了保证只有设备拥有者才能使用这个管理页面工具，该设备通常也设有登录界面，只有输入正确的用户名和密码的用户才能进入管理页面。然而在设备出售时，制造商给每一个型号的设备提供的默认用户名和密码都是一样的，很多用户购买这些设备回来之后都不会去修改设备的默认的用户名和密码，这就使得黑客们有机可乘。他们只要通过简单的扫描工具就能很容易地找出这些设备的地址并尝试用默认的用户名和密码去登录管理页面，如果成功则立即取得该路由器的控制权。此外，大多数路由器都有远程管理特性，允许用户从网络外部进入系统并实施管理，因此若无特别需要应避免使用远程管理。

3. 无线网络数据加密

黑客攻击计算机只要在无线路由器/中继器的有效范围内的话，就有很大机会访问到该无线网络，一旦它能够访问该内部网络，那么该网络中所有传输的数据对他来说都是透明的。如果这些数据都没经过加密的话，黑客就可以通过一些数据包嗅探工具来抓包，分析并窥探到其中的隐私。如果开启无线网络加密，这样即使用户在无线网络上传输的数据被截取了也没那么容易被解读。用户可以采用无线对等协议（WEP）或 Wi-Fi Protected Access（WPA）方式加密无线传输数据，这样可以实现类似有线数据传输的保护。另外，最好能做到定期更换密钥，可考虑每个季度更换一次密钥，这样安全性会大幅度提高。

4. 采用 MAC 地址过滤

介质访问控制（Media Access Control，MAC）地址是厂商生产的网卡的地址，对于每块无线网卡都拥有一个唯一的 MAC 地址，为 AP 设置基于 MAC 地址的访问控制表，确保只有经过注册的设备才能进入网络。如果某个计算机的 MAC 地址没有出现在该列表上，它就无法连接到用户的路由器和网络。无线 MAC 过滤可以让无线网络获得较高的安全性，但这并非绝对有效的安全方法。经验老到的黑客可以为自己的计算机设定一个虚假的 MAC 地址，但他们需要知道用户的授权计算机列表上有哪些 MAC 地址。遗憾的是，因为 MAC 地址在传输时没有经过加密，所以黑客只要探测或监控到网络上传输的数据包，就能知道列表上有哪些 MAC 地址。所以 MAC 地址过滤只能对付黑客新手。

5. 禁止 SSID 广播

服务区标识符（Service Set Identifier，SSID）是无线局域网用于身份认证的登录名。在无线网络中，各路由设备有个很重要的功能，那就是服务区标识符广播，即 SSID 广播。最初这个功能主要是为那些无线网络客户端流动量特别大的商业无线网络而设计的。开启 SSID 广播利于无线网络客户端自动接收 SSID 号，从而利用这个 SSID 号连接到无线网络。但是这个功能存在极大的安全隐患，它自动地为想进入该网络的黑客打开了门户。在商业网络里，由于为了满足经常变动的无线网络接入端，必定要牺牲安全性来开启这项功能，但是对于家庭或拥有固定客户端的用户来讲，完全没有必要开启这项功能。出于安全考虑，可以将多数无线 AP 或无线路由器在出厂时默认的“允许广播 SSID”设为“不广播 SSID”，此时用户就要手工设置 SSID 才能进入相应的网络。

6. 有效管理 IP 分配

分配 IP 地址有 DHCP 分配和静态地址分配两种方式，判断无线网络使用哪一种分配

IP 的方式对网络的安全至关重要。为了方便客户端设置，一般来说无线路由器都会带有 DHCP 功能。DHCP 分配可以降低繁重的管理工作，比较适用于公共场所，但使用 DHCP 为网络中的客户端动态分配 IP 将导致另外一个安全隐患：网络管理员不太容易分辨出在线用户中哪些是合法用户，哪些是非法用户。如果关闭 DHCP，那么网络管理员只需查看当前在线用户，如果发现其 IP 地址不是分配的，则说明是非法侵入的。采用静态地址可以对局域网中计算机的 IP 地址进行控制，能有效地防止黑客自动获取。

7. 使用 VPN

使用虚拟专用网络(VPN)可以在一个不安全的公共网络中建立一个安全的、加密的网络。VPN 的特点是：对两个点或两个网络之间的通信进行加密通常是基于软件的，可以提供不同的加密级别。VPN 的优点在于：在公司各地的办事处之间实现安全和便利的通信；为移动员工提供廉价的网络访问途径；为在家上网的用户提供安全的网络访问。

VPN 通过以下两种方式提供安全加密的通信：

(1) 用户到网络，即远程访问模型。采用这种模式，远程客户端通过一个公共网络(例如 Internet)来连接，使远程用户成为公司网络的一部分。

(2) 网络到网络，即站点到站点模型。采用这种模式，分支机构的网络可通过一个公共网络(例如 Internet)连接另一个分支机构的网络，避免铺设一个昂贵的广域网。

在 WEP 加密基础上再使用 VPN 加密，黑客就必须对数据进行两次解密，第一次是对比较容易破解的 WEP 解密，第二次是对比较可靠的 VPN 解密，但黑客无法轻松获得 VPN 密码、证书或智能卡密钥，故成功破解 VPN 的概率非常低。但同时使用 WEP 和 VPN 会带来额外的系统开销。此外，在无线网络中使用 VPN 时需要在每个用户的设备上安装客户端软件。

8. RADIUS

RADIUS 协议负责对系统的远程连接进行身份认证，为网络资源提供授权及记录日志。VPN 和 WLAN 都提供对它的支持，用它可控制与用户连接的几乎每一个方面。

RADIUS 的主要功能：

(1) 身份认证：通过一个中心安全数据库对任何一个远程或 WLAN 用户的用户名和密码进行校验，确保只有合法身份的用户才能访问网络。

(2) 授权：对于每个新连接都为远程访问或 WLAN 用户访问点设备提供相应的信息。

(3) 日志记录：记录所有远程和 WLAN 连接，包括用户名和连接持续时间。

RADIUS 除了用户身份认证，还可应用于访问点身份认证，即实现双向身份认证，这样黑客就无法假冒访问点对网络进行攻击。RADIUS 可全面地控制访问授权，例如时间限制及密钥重新生成周期。

10.3.4　无线网络安全的管理机制

无线网络的安全取决于其安全措施中最薄弱的环节，因此除了加强技术手段外，还应进行合理的物理布局及实施严格的管理。

1. 合理进行物理布局

进行网络布局时要考虑两方面的问题：一是限制信号的覆盖范围在指定范围内，二是

保证在指定范围内的用户获得最佳信号。这样入侵者在范围外将搜寻不到信号,或者只能搜寻到微弱信号,不利于进行下一步的攻击行为。因此,合理确定接入点的数量及位置是十分重要的,既要让其具有充分的覆盖范围,又要尽量避免无线信号受到其他无线电的干扰而减小覆盖范围或减弱信号强度。

2. 建立安全管理机制

无线网络信号在空气中传播,也就注定了它更脆弱、更易受到威胁,因此建立健全的网络安全管理制度是尤为重要的。这应明确网络管理员和网络用户的职责和权限,在网络可能受到威胁或正在面临威胁时能及时检测、报警,在入侵行为得逞时能提供资料、依据及应急措施,以恢复网络正常运行。

3. 加强用户安全意识

现在的无线设备比较便宜,而且安装简单,如果网内的用户私自安装无线设备,他们往往只采取有限的安全措施,这样极有可能将网络的覆盖范围超出可控范围,将内部网络暴露给攻击者。而这些用户通常也没有意识到私自安装接入点带来的危险,因此必然要让用户清楚自己的行为可能会给整个网络带来的安全隐患,加强网络安全教育,提高用户的安全意识。

10.4 IEEE 802.11 的安全性

10.4.1 IEEE 802.11 概述

作为全球公认的局域网权威,IEEE 802 工作组建立的标准在过去 20 多年里在局域网领域内独领风骚。这些协议包括 IEEE 802.3 Ethernet 协议、802.5 Token Ring 协议、802.3z 100BASE-T 快速以太网协议。在 1997 年,经过了 7 年的工作以后,IEEE 发布了 802.11 协议,这也是在无线局域网领域内的第一个国际上被认可的协议。在 1999 年 9 月,他们又提出了 802.11bHigh Rate 协议,用来对 802.11 协议进行补充,802.11b 在 802.11 的 1Mbps 和 2Mbps 速率下又增加了 5.5Mbps 和 11Mbps 两个新的网络传输速率。

和其他 IEEE 802 标准一样,802.11 协议主要工作在 ISO 协议的最低两层上,也就是物理层和数字链路层。任何局域网的应用程序、网络操作系统或者像 TCP/IP、Novell NetWare 都能够在 802.11 协议上兼容运行,就像他们运行在 802.3 Ethernet 上一样。IEEE 802.11 协议的实现层次如图 10-7 所示。

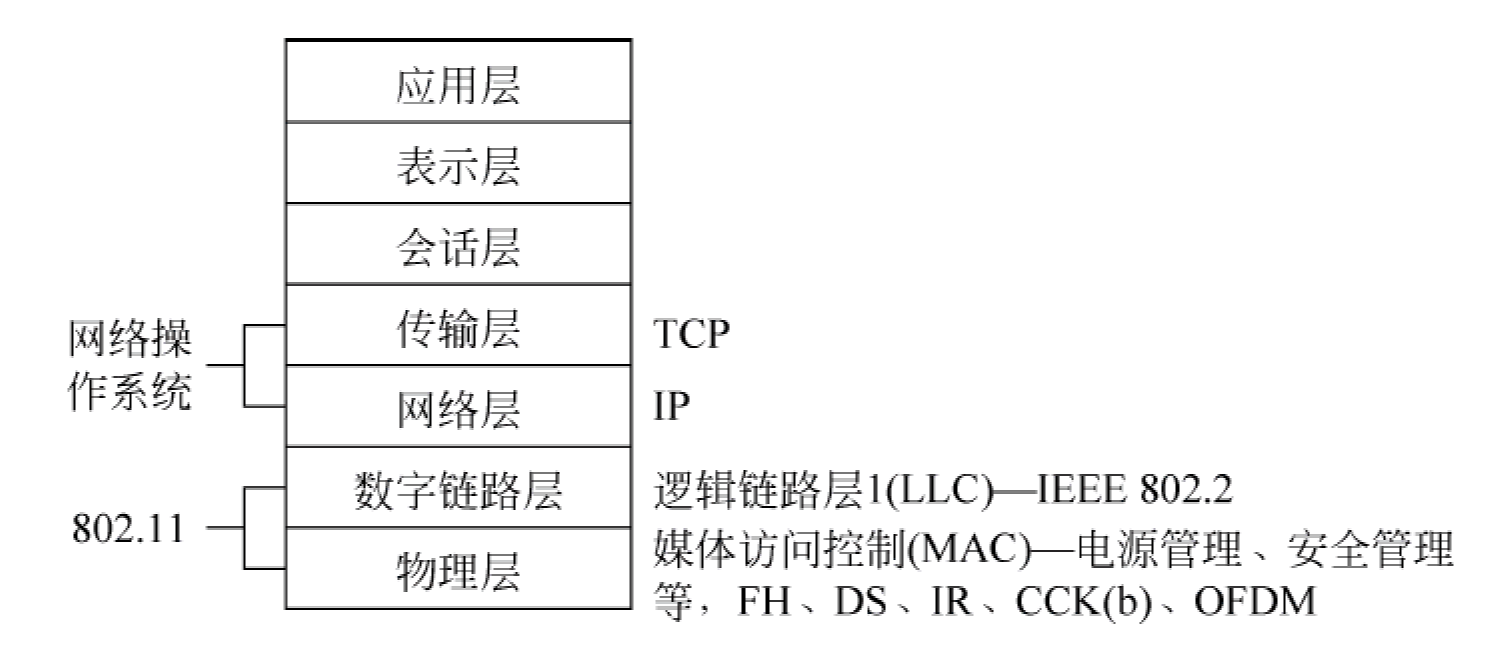


图 10-7 IEEE 802.11 协议的实现层次

1. IEEE 802.11 物理层

在 IEEE 802.11 物理层主要定义了红外线（Infrared，IR）、直接序列扩频（DS）和跳频扩频（FH）3 种传输技术。

（1）红外线传输技术

采用接近可见光的 850～950nm 信号，无须对准，依靠反射和直视红外能量进行通信。红外辐射不能穿透墙壁，穿过窗户时也有显著衰减。这种特性使 IR 仅限于单个物理房间中。使用 IR 的多个不同局域网可在仅有一墙之隔的相邻房间中毫无干扰地工作，且不存在被窃听的可能。IR 传输一般采用基带传输方案，主要是脉冲调制方式。IR 定义了两种调制方式和数据速率：基本接入速率和增强接入速率。基本接入速率是基于 1Mbps 的 16 PPM（脉冲位置调制）调制；增强接入速率是基于 2Mbps 的 4PPM 调制。

（2）直接序列扩频技术

把要传输的信息直接由高码速的扩频码序列编码后，对载波进行伪随机的相位调制，以扩展信号的频谱。而在接收端，用相同的扩频码序列进行解扩，把展宽的扩频信号还原成原始信息。在扩频传输中用得最多的扩频码序列是伪噪声码序列，它具有伪随机的特点。DS 采用差分二进制相移键控（DBPSK）和差分四进制相移键控（DQPSK）来分别提供 1Mbps 和 2Mbps 的数据传输速率。

（3）跳频扩频技术

它是用伪随机码序列去进行频移键控调制（FSK），使载波工作的中心频率不断地、随机地跳跃改变，而干扰信号的中心频率却不会改变。只要收、发信机之间按照固定的数字算法产生相同的伪随机码，就可以把调频信号还原成原始信息。FH 也有 1Mbps 和 2Mbps 两种数据传输速率，前者采用二值的高斯频移键控（2GFSK），后者采用四相高斯频移键控（4GFSK）。

2. IEEE 802.11 数据链路层

IEEE 802.11 的数据链路层由两个子层构成，逻辑链路层 LLC（Logic Link Control）和媒体控制层 MAC（Media Access Control）。802.11 使用和 802.2 完全相同的 LLC 子层和 802 协议中的 48 位 MAC 地址，这使得无线和有线之间的桥接非常方便，但是 MAC 地址只对无线局域网唯一。

IEEE 802.11 无线媒体访问协议称为基于分布方式的无线媒体访问控制协议，它支持自组织结构（Ad-hoc）和基础结构（Infrastructure）两种类型的 WLAN。它有两种方式，即分布协调功能（Distributed Coordination Function，DCF）和点协调功能（Point Coordination Function，PCF）。

1）分布协调功能

DCF 是 IEEE 802.11 最基本的媒体访问方法，其核心是 CSMA /CA。它包括载波检测机制、帧间隔和随机退避规程。DCF 在所有站点（Station，STA）上都进行实现，用于 Ad-hoc 和 Infrastructure 网络结构中，提供争用服务。DCF 有两种工作方式：基本工作方式，即 CSMA/CA 方式和 RTS/CTS 方式。CSMA/CA 是基础，RTS/CTS 只是 CSMA/CA 之上的可选机制。

2）点协调功能

PCF是可选的媒体访问方法，用于Infrastructure网络结构中。它使用集中控制的接入算法，一般在接入点AP实现集中控制，用类似轮询的方法将发送数据权轮流交给各个站，从而避免了碰撞的产生。对于时间敏感的业务，例如分组语音，就应使用提供无争用服务的点协调功能PCF。

10.4.2 IEEE 802.11的认证服务

IEEE 802.11提供了两种类型的认证服务：开放系统认证和共享密钥认证。认证类型由认证管理帧的帧体指出，因此认证帧能自己识别认证算法。所有认证类型的管理帧应该是单播，因为认证是在对等的工作站间进行(不允许广播认证)。管理帧中的解除认证帧是报告性的，因此可以作为组地址帧发送。

两个工作站间在进行了一次成功的认证信息交换后即存在了互相认证关系。认证可以在一个基本服务集BSS中的工作站和AP间进行，也可以用在一个独立基本服务集IBSS中的两个工作站间进行。

1. 开放系统认证

开放系统认证是可用认证算法中简单的一种，本质上是一个空的认证算法。如果接收方的认证类型设为开放系统认证，那么用此算法请求认证的任何类型的工作站将成为已认证。但是开放系统认证请求不能保证一定会成功，因为一个工作站可以拒绝一些类型的工作站。开放系统认证是默认的认证算法。

开放系统认证由两步来完成：第一步用自己的标识请求认证；第二步是认证结果，如果结果是成功，那么工作站已相互认证了。

2. 共享密钥认证

共享密钥认证既支持知道共享密钥的工作站间的认证，也支持不知道共享密钥的工作站间的认证。IEEE 802.11共享密钥认证不需要明文传输密钥，它使用WEP加密机制。因此，共享密钥认证方案只有采用WEP选项时才能使用。另外，共享密钥认证算法也用于使用WEP作为认证算法的工作站。

假定共享密钥是通过一个独立于IEEE 802.11的安全渠道传输到指定工作站的，经过MAC管理路径，共享密钥包含在一个只写的MIB属性里，属性的只读保证密钥值仍旧是MAC内部的。

在共享密钥认证信息交换过程中既传输不加密信息也传输加密信息，这有利于让授权的用户发现用于密钥/初始化向量对交换的伪随机序列。因此在实现中要避免在后续帧中使用同一密钥/初始化向量对。一个工作站只有它的加密实现选项的属性为“真”时才进行共享密钥认证。

10.4.3 IEEE 802.11的保密机制

1. WLAN ESSID

首先在每一个接入点(Access Point)内都会写入一个服务区域认证ID(WLAN ESSID)，每当端点要连上AP时，AP会检查其ESSID是否与其相同，如果不符就拒绝服务。

2. Access Control Lists

也可以将无线局域网络只设定为给特定的节点使用,因为每一张无线网卡都有一个唯一的 MAC Address,只要将其分别输入 AP 即可。相反地,如果有网卡被偷或发觉有存取行为异样,也可以将这些 MAC Address 输入,禁止其再次使用。利用这个存取控制机制,如果有外来的不速之客得知公司使用的 WLAN ESSID 也一样会被拒绝在外。

3. Layer 2 Encryption

IEEE 802.11b WEP 采用对称性加密算法 RC4,在加密与解密端均使用 40b 长度的相同密钥(Secret Key)。这个密钥被输入每一个客户端和接入点之中。而所有资料的传输与接收,不管在客户端或存取端,都使用这个共享密钥(Share Key)来做加密与解密。WEP 也提供客户端使用者的认证功能,当加密机制功能启用,客户端要尝试连接上接入点时,接入点会核发出一个测验挑战值封包(Challenge Packet)给客户端,客户端再利用共享密钥将此值加密后送回接入点以进行认证比对,如果无误才能获准存取网络的资源。

10.4.4 IEEE 802.11b 安全机制的缺点

纵使 IEEE 802.11b 标准能提供完整的保密机制给无线局域网络使用,却仍然存在着以下缺点。

1. 无线局域网络 ESSID 的安全性

利用特定接入点的 ESSID 来做存取控制,照理说是一个不错的保护机制,它强制每一个客户端都必须要有跟接入点相同的 ESSID 值。但是,如果在无线网卡上设定其 ESSID 为 ANY 时,它就可以自动地搜寻在信号范围内所有的接入点,并试图连接上它。

2. WEP 的安全性

IEEE802.11 的 WEP 是为了克服无线信号的易受窃听攻击而设计的协议,所以 WEP 在安全上较弱,存在漏洞。WEP 提供 40b 长度的加密密钥,对于一般的黑客尚足以防范,但如果有专业的网络黑客刻意地要偷听及窃取用户数据传输期间的私密资料却是易如反掌。40b 的长度可以排列出 2^{40} 的 Keys,而现今 RSA 破解的速度可每秒尝试破解出 2.45×10^{9} 的 Keys,也就是说 40b 长度的加密资料在 5min 之内就可以被破解出来。所以各家网络厂商便推出 128b 长度的加密密钥。

ISAAC(Internet Security Applications Authentication and CryptographyGroup)列出了 4 种攻击 WEP 的方法:

(1) 截获 WEP 数据流,主要是通过分析明文和密文的对应关系。

(2) 主动攻击,例如插入非法的数据包。

(3) 主动攻击并获取 WEP 的数据内容和上层报文的头信息(如 IP 地址等)。

(4) 基于表的攻击,通过截获并记录 IV,可以推出 RC4 算法的密钥信息。

3. 用户身份认证方法的缺陷

IEEE 802.11 规定的开放系统认证机制中任何移动接入都可以加入 BSS,并可以与 AP 通信,能听到所有未加密的数据,可见这种方法不存在安全性。共享密钥认证是一种请求响应认证机制,能提供较高的安全系数。但攻击者易获得 WEP 加密前后的询问信息,将二者

进行异或运算就可以得到密钥序列，从而冒充合法身份介入 WLAN。

此外 IEEE 802.11 还缺少一种双向认证机制，接入点可以验证客户机的身份，而客户机不能验证接入点的身份。如果一个虚假接入点被放置在无线局域网中，它可以通过“劫持”合法客户机成为拒绝访问的平台。

10.5 蓝牙安全

蓝牙技术提供了一种短距离的无线通信标准，它的无线传输特性使其非常容易受到攻击，因此安全机制在蓝牙技术中显得尤为重要。

10.5.1 蓝牙技术概述

蓝牙(Bluetooth)无线接入技术发布于 1998 年，“蓝牙”原是 10 世纪统一了丹麦的国王的名字，现取其“统一”的含义，用来命名意在统一无线局域网通信标准的蓝牙技术。蓝牙作为近距离无线连接技术是 3G 和 IEEE 802.11 系列的补充。蓝牙无线数字传输标准是由爱利信、IBM、Intel、诺基亚和东芝等五大 IT 业著名公司共同提出的。

蓝牙的目标是实现无线数据和语音传输的开放式标准，用微波取代传统网络中错综复杂的电缆，将各种通信设备、计算机及其终端设备、各种数字数据系统甚至家用电器采用无线方式连接起来，以进行方便快捷、灵活安全、低成本低功耗的数据和语音通信。它的传输距离为 10cm～10m，如果增加功率或是加上某些外设便可达到 100m 以上的传输距离。

两个蓝牙设备在传输数据前要进行鉴权以确认身份，而鉴权过程要用到两个蓝牙设备的公共链路字。若两个蓝牙设备不是第一次通信，那么它们就会使用各自以前存储的公共链路字。若两个蓝牙设备是第一次通信，就需要在两个蓝牙设备上分别生成初始字暂时作为公共链路字来进行鉴权，随后两个蓝牙设备协商公共链路字并分别存储，以用于下一次通信。公共链路字可以采用其中一个蓝牙设备的单元字，也可以采用两个蓝牙设备的组合字。两个蓝牙设备在鉴权之后就可以进行加密传输数据了。整个流程如图 10-8 所示。

10.5.2 蓝牙技术特点

1. 无线性

蓝牙技术最初是以取消连接各种电器之间的连线为目标的。蓝牙技术主要面向网络中的各种数据及语音设备，例如 PC、拨号网络、笔记本计算机、打印机、传真机、数码相机、移动电话、高品质耳机等。蓝牙通过无线的方式将它们连成一个围绕个人的网络，省去了用户接线的烦恼，在各种便携设备之间实现资源共享。

2. 开放性

与生俱来的开放性赋予了蓝牙强大的生命力。从它诞生之日起，蓝牙就是一个由厂商们自己发起的技术协议，完全公开，而并非某一家独有和保密。只要是 SIG 的成员，都有权无偿使用蓝牙的新技术。而且蓝牙技术标准制订后，任何厂商都可以无偿地拿来生产产品，只要产品通过 SIG 组织的测试并符合蓝牙标准后，即可投入市场。

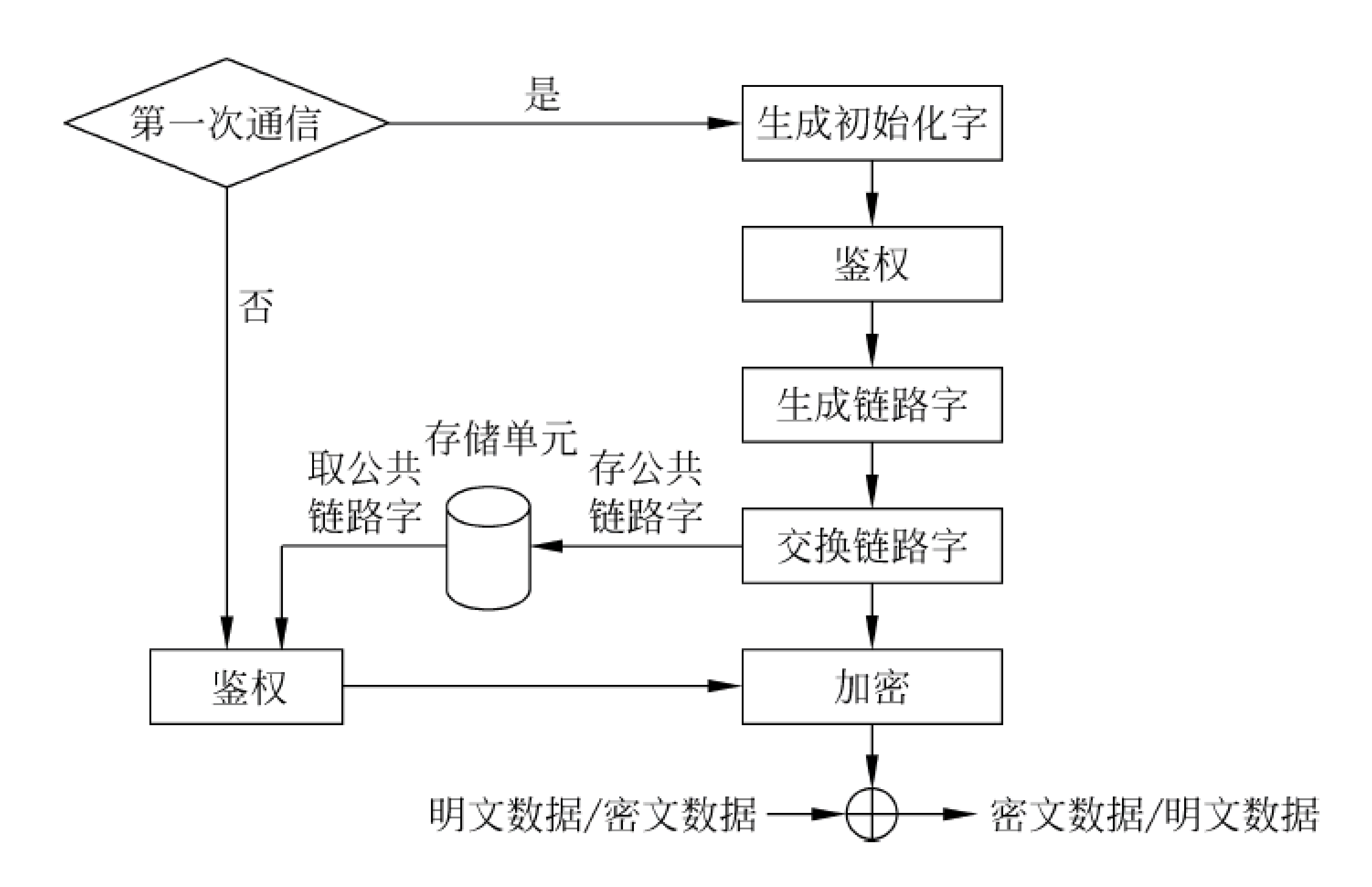


图 10-8　两个蓝牙设备通信的流程

3. 蓝牙产品的互操作性和兼容性

蓝牙产品在满足蓝牙规范的前提下,还必须通过 SIG 的认证程序才能走向市场。这就保证了即使是不同公司的蓝牙产品,也可实现互操作和数据共享,达到完全兼容的目的。

4. 对人体安全影响不大

随着无线技术的深入人心,辐射也成了消费者非常关心的问题。由世界卫生组织、IEEE 等专家组成的小组表示,检测中并未发现蓝牙产品的辐射对人体有影响。蓝牙产品的输出功率仅为 1mW。是微波炉使用功率的百万分之一,而且这些输出中只有一小部分被人体吸收。

10.5.3　蓝牙系统安全性参数

(1) 蓝牙设备地址(BD_ADDR):每个蓝牙设备都有一个全球唯一的 48 位 IEEE 802 标准的地址。蓝牙设备地址是公开的,可以通过人机接口交互或蓝牙设备的查询规则取得。蓝牙地址分 3 个域:LAP 域(24 位低地址部分)、UAP 域(8 位高地址部分)和 NAP 域(16 位重要地址部分)。LAP 和 UAP 构成蓝牙设备地址的重要部分。蓝牙设备地址可获得的整个地址空间为 2^{32}。

(2) 个人确认码(Personal Identification Number,PIN):由蓝牙单元提供的 1～16 位(八进制)数字,可以固定或者由用户选择。一般,这个 PIN 码是随单元一起提供的一个固定数字。但当该单元有人机接口时,用户可以任意选择 PIN 的值,从而进入通信单元。蓝牙基带标准中要求 PIN 的值是可以改变的。

(3) 鉴权字:是长度为 128b 的数字,用于系统的鉴权。

(4) 加密字:长度 8～128b,可以改变。这是因为不同的国家有许多不同的对加密算法的要求,同时也是各种不同应用的需要,还有利于算法和加密硬件系统的升级。

10.5.4　蓝牙采用的安全技术

蓝牙技术标准除了采用跳频扩频技术和低发射功率等常规安全技术外,还采用内置的

安全机制来保证无线传输的安全性。

1. 安全模式

在蓝牙技术标准中定义了3种安全模式：

(1) 安全模式1：又称无安全模式，在该模式下，蓝牙设备可以屏蔽所有的安全机制。

(2) 安全模式2：又称服务级安全模式，该模式为应用程序提供多种灵活的访问策略。

(3) 安全模式3：又称链路级安全模式，该模式要求设备在建立链路连接前要启动链路级安全措施。

安全模式2与安全模式3的本质区别在于：安全模式2下的蓝牙设备在信道建立以后启动安全性过程，即其安全性过程在较高协议进行；安全模3下的蓝牙设备在信道建立以前启动安全性过程，即其安全性过程在低层协议进行。

2. 设备和业务的安全等级

蓝牙技术标准为蓝牙设备和业务定义安全等级，其中设备定义了3个级别的信任等级：

(1) 可信任设备：设备已通过鉴权，存储了链路密钥，在设备数据库中标识为“可信任”。可信任设备可以无限制地访问所有的业务。

(2) 不可信任设备：设备已通过鉴权，存储了链路密钥，但在设备数据库中没有标识为“可信任”。不可信任设备访问业务是受限的。

(3) 未知设备：无此设备的安全性信息，为不可信任设备。

对于业务，蓝牙技术标准定义了3种安全级别：需要授权与鉴权的业务、仅需鉴权的业务以及对所有设备开放的业务。一个业务的安全等级由下述3个属性决定，它们保存在业务数据库中。

(1) 需授权：只允许信任设备自动访问的业务。不信任的设备需要在授权过程完成后才能访问该业务。授权总是需要鉴权以确认远端设备是正确的设备。

(2) 需鉴权：在连接到应用程序之前，远端设备必须接受鉴权。

(3) 需加密：在允许访问业务前必须切换到加密模式下。

3. 链路级安全参数

蓝牙技术在应用层和链路层上提供了安全措施。链路层采用如表10-3所示的4种不同实体来保证安全。所有链路级的安全功能都是基于链路密钥的概念实现的，链路密钥是对应每一对设备单独存储的一些128b的随机序列。

表10-3 链路层用于鉴权和加密的实体

通 用 性	支持客户机/服务器模式
单元密钥 K_A	在安装蓝牙设备时由单元A产生
联合密钥 K_{AB}	由单元A和B产生，每一对设备有各自的联合密钥，其在需要更多的安全性时使用
临时密钥 K_{master}	此密钥在主设备需同时向多个从设备传输数据时使用，在此次会话过程中它将临时替代原有的链路密钥
初始化密钥 K_{init}	在初始化过程中使用，用于保护初始化参数的传输

4. 密钥管理

蓝牙系统用于确保安全传输的密钥有几种，其中最重要的密钥是用于两个蓝牙设备之

间鉴权的链路密钥。加密密钥可以由链路密钥推算出来，这将确保数据包的安全，而且每次传输都会重新生成。最后还有 PIN 码，用于设备之间互相识别。

5. 加密算法

蓝牙系统加密算法为数据包中的净荷(即数据部分)加密，其核心部分是数据流密码机 E0，它包括净荷密钥生成器、密钥流生成器和加/解密模块。由于密钥长度从 8～128b 不等，信息交互双方必须通过协商确定密钥长度。

有几种加密模式可供使用，如果使用了单元密钥或者联合密钥，广播的数据流将不进行加密。点对点的数据流可以加密也可以不加密。如果使用了主密钥，则有 3 种可能的模式：

(1) 加密模式 1：不对任何数据进行加密。

(2) 加密模式 2：广播数据流不加密，点对点数据流用临时密钥 K_{master} 进行加密。

(3) 加密模式 3：所有数据流均用临时密钥 K_{master} 进行加密。

每个应用程序对密钥长度有严格的限制，当应用程序发现协商后得到的密钥长度与程序要求不符，就会废弃协商的密钥长度。这主要是为了防止恶意用户通过协商过程减小应用程序密钥长度，以便对系统造成破坏。

6. 认证机制

两个设备第一次通信时，借助“结对”初始化过程生成一个共用的链路密钥，结对过程要求用户输入 16B 的 PIN 到两个设备，根据蓝牙技术标准，结对过程如下：根据用户输入的 PIN 生成一个共用随机序列作为初始化密钥，此密钥只用一次，然后即被丢弃。在整个鉴权过程中，始终检查 PIN 是否与结对设备相符。生成一个普通的 128b 随机序列链路密钥，暂时储存在结对的设备中。只要该链路密钥储存在双方设备中就不再需要重复结对过程，只需实现鉴权过程。基带连接加密不需要用户的输入，当成功鉴权并检索到当前链路密钥后，链路密钥会为每个通信会话生成一个新的加密密钥，加密密钥长度依据对安全等级而定，一般在 8～128b，最大的加密长度受硬件能力的限制。

为防止非授权用户的攻击，蓝牙标准规定，如果认证失败，蓝牙设备会推迟一段时间重新请求认证，每增加一次认证请求，推迟时间就会增加一倍，直到推迟时间达到最大值。同样认证请求成功后，推迟时间也相应地成倍递减，直到达到最小值。

10.5.5　蓝牙安全技术存在的问题

1. 隐私问题

由于蓝牙设备内的蓝牙地址具有全球唯一性，一方面保证了设备不会被人冒用，而 BD-ADDR 的唯一性也导致了用户隐私问题。用户在移动使用的过程中很容易被人追踪，个人行为容易暴露，私密性可能会受到侵害。

2. PIN 问题

PIN 是唯一的可信的用于生成密钥的数据，链路密钥和加密密钥都与它有关。为了初始化一个安全连接，两个蓝牙设备必须输入相同的 PIN 码。用户有可能将其存在设备上，或者输入过于简单，所以 PIN 易受到攻击，解决的方法是使用较长的 PIN，或者使用密钥变更系统。

3. 链路密钥

鉴权和加密都是基于双方共享的链路密钥,所有在程序中的其他信息都是公开的。这样,某一设备很可能利用早就得到链路密钥以及一个伪蓝牙地址计算出加密密钥,从而监听数据流。假设设备A、B使用A的单元密钥作为链路密钥,与此同时或者是稍后设备C也使用设备A的单元密钥与A进行通信;B由于已经获得A的单元密钥,就可以使用该密钥和一个伪造的BD-ADDR计算出加密密钥,监听信息流。并且可以把自己当成A来应付C的鉴权或是当成C应付A的鉴权。

思 考 题

(1) 什么是无线网络?简单描述无线网络与有线网络的差别。
(2) 无线网络按区域来分可以分为哪几类?
(3) 目前无线网络所采用的技术标准主要有哪些?
(4) 请列举出无线网络的主要设备并对其功能进行简要介绍。
(5) 简要说明无线网络的安全技术。
(6) 请列举出无线网络受到的主要攻击及其防治措施。
(7) 简述一下IEEE 802.11的物理层所采用的技术。
(8) IEEE 802.11提供的认证服务有哪几种?
(9) 简述一下两个蓝牙设备进行通信的流程。
(10) 蓝牙技术的特点有哪些?
(11) 目前蓝牙技术存在哪些主要问题?

参 考 文 献

[1] 戴红,王海泉,黄坚.计算机网络安全.北京:电子工业出版社,2004.
[2] 金纯.无线网络安全:技术与策略.北京:电子工业出版社,2004.
[3] 程民利.无线网络中的安全技术研究.东南大学硕士学位论文,2005.
[4] 徐艳.蓝牙技术安全性分析与安全策略.华北科技学院学报,2005,2(4):42~45.
[5] 李吉平,郭凤宇,姜春.浅谈无线网络的安全隐患及应对措施.科技信息,2009,(3):66~67.
[6] 谭瑛.无线网络环境下的网络安全.现代计算机,2010,(1):142~144.
[7] 王曼珠,何文才,杨亚涛,魏占祯.无线局域网IEEE 802.11的安全缺陷分析.微电子学与计算机,2005,22(7).
[8] 李浩,高泽华,高峰,赵荣华.IEEE 802.11无线局域网标准研究.计算机应用研究,2009,26(5):1617~1620.
[9] 张俊.蓝牙技术及其安全机制研究.电脑知识与技术,2008,3(9):1941~1942.
[10] 尤麦峰,梁计春,等.直接序列扩频和调频扩频.装甲兵工程学院学报,2001,15(1).
[11] Dr. Cyrus Peikari,Seth Fogie著.无线网络安全.周靖译.北京:电子工业出版社,2004.

第 11 章　恶意软件攻击与防治

本章学习目标

随着信息化的飞速发展，网络已经成为人们交流的主要途径。然而网络在传播一些先进工具与技术的同时，也开始出现了恶意软件，影响人们的正常使用，甚至造成重大损失。因此加强对恶意软件检测技术的研究，对维护网络安全，推动我国 Internet 产业的持续、稳定、快速、健康发展以及保护广大网民的合法权益都具有重要意义。

通过对本章的学习，应掌握以下内容：

(1) 了解恶意软件的基本知识。

(2) 掌握常见恶意软件的相关危害及防治方法。

(3) 掌握典型恶意软件的攻防方法。

当今的网络威胁比以往都更具敌意。在仿冒和垃圾邮件方面取得的新进展说明攻击者的手段已经更趋向于心理学方面而非技术方面。通过电子邮件和 Web，用户成了攻击的目标。仿冒网站看上去如此可信，使得许多人没办法看出与真实网站的不同，从而交出了自己的敏感信息，例如网上银行的用户名和密码。根据 McAfee 的网站指南，在他们所做的间谍软件调查问卷（询问受访人一个网站是否安全）中，12 万名受访者中的 95％错误地认为一个含有恶意软件的网站是安全的。

目前，蠕虫和病毒对个人用户和公司网络的威胁已经显著减少了。但是，病毒爆发的停止不是因为病毒编写者决定洗手不干，而是因为他们的主要目标——公众注意力，已经不再让他们感兴趣了。病毒编写者想要得到的是金钱、用户的敏感信息以及对未授权系统的持续访问。因此他们改变了方法、技术和工具，变得更加谨慎和针对特定目标。

Microsoft 操作系统远程漏洞的减少和边界安全产品的广泛使用迫使攻击者提升自己的水平。

11.1　恶意软件的基本概念

11.1.1　什么是恶意软件

根据中国互联网协会的初步定义，恶意软件是指在未明确提示用户或未经用户许可的情况下，在用户计算机或其他终端上安装运行侵犯用户合法权益的软件，但已被我国现有法律法规规定的计算机病毒除外。

1981 年 Richard Skrenta 针对 Apple Ⅱ系统编写的 Elk Cloner 是历史上第一个大规模爆发的恶意程序。2010 年 5 月出现的 Gilda 蠕虫会影响运行 Windows 7 和 Windows Server 2008 R2 的计算机，它绕过或禁止用户账户控制，删除内置显示驱动文件而造成蓝屏宕机。经过 30 多年的发展，恶意软件数量已经超过几百万种，包括广告软件、僵尸网络、计

算机病毒、计算机蠕虫、间谍软件、Rootkits、特洛伊木马等13大类。

一个更为重要的趋势是，随着计算机犯罪赢利组织的出现，恶意软件功能日益复杂化，其质量也愈来愈高。大多数恶意软件目的是借助窃取信息、破坏数据、危害系统以达到更高目标。恶意软件危害的根源在于用户在执行他们并不了解的程序。多数用户不知道程序会做什么，也没有可靠的方法去发现程序的行为。有安全概念的用户在执行未知程序前会使用病毒扫描工具。尽管现代病毒扫描器具有扩展经验数据的能力，但是恶意软件检测的主要方法仍然是简单的模式匹配机制，即把未知文件与已有的恶意软件特征码数据库进行比较，对于针对某个组织专门定制的恶意软件，既不会广泛传播，也不会送到反病毒厂家，这种方法显然无能为力。

恶意软件有八大特征：

(1) 强制安装。未明确提示用户或未经用户许可，在用户计算机或其他终端上安装软件。

(2) 难以卸载。未提供通用卸载方式，或在不受其他软件影响、人为破坏的情况下，卸载后仍然有活动程序。

(3) 浏览器劫持。未经用户许可修改用户浏览器或其他相关设置，迫使用户访问特定网站或导致用户无法正常上网。

(4) 广告弹出。在未明确提示用户或未经用户许可的情况下，利用安装在用户计算机或其他终端上的软件弹出广告。

(5) 恶意收集用户信息。未明确提示用户或未经用户许可，恶意收集用户信息。

(6) 恶意卸载。未明确提示用户、未经用户许可，或误导、欺骗用户卸载非恶意软件。

(7) 恶意捆绑。在软件中捆绑已被认定的恶意软件。

(8) 其他。侵犯用户知情权、选择权的恶意行为。

11.1.2 恶意软件的分类

恶意软件的种类较多，例如Virus(病毒)、Worm(蠕虫)、Bot(蠕虫)、Trojan Horse(特洛伊木马)、Exploit(漏洞利用程序)、Backdoor(后门)、Rootkit、Spyware(间谍软件)、Spamware(垃圾信息发送软件)、Adware(垃圾广告软件)等。

恶意软件有多种分类方法，对反病毒厂商而言最典型的方法是通过其意图(例如特洛伊木马、蠕虫、邮件携带者等)和危害程度(潜在损失、爆发可能性、实际爆发报告等)加以分类。这些因素可以用于生成整体的风险评价。对最终用户而言，恶意软件通常是指那些他们未请求或不需要的、危害他们计算机系统的软件。

恶意软件会暗中安装，并秘密寻求最大利益化和可重用性。在现代计算机和Internet时代，犯罪组织大量地使用恶意软件。入侵目标系统并达到特定目的的完整过程包含3个阶段：获取对目标系统的远程控制权；维持对目标系统的远程控制权；通过远程控制在目标系统上完成特定业务逻辑。这里所说的目标系统可以是目标主机系统，也可以是目标网络设备。

根据不同恶意软件所完成的功能在完整的入侵过程中所处阶段的不同，将恶意软件分为以下3种类型。

1. 获取目标系统远程控制权类

获取目标系统远程控制权类的恶意软件的基本特征是在未授权的条件下，利用各种手段获取对目标系统的远程控制权的功能，即具有完整入侵过程中第一阶段的功能。

1）Exploit（漏洞利用程序）

它是第一类恶意软件的基本形式。Exploit 利用操作系统或应用程序中存在的缺陷(bug)可达到以非授权的方式远程控制目标系统或提升本地用户权限的目的。具体过程为：构造特定的输入数据并提交给存在缺陷的操作系统程序或应用程序，使这些有缺陷的程序在所构造的输入数据下改变正常的程序流程，从而导致未授权用户可以远程控制目标系统，或使本地用户获得更高的权限。

典型的 Exploit 类恶意软件有 Buffer Overflow Exploit（缓冲区溢出漏洞利用工具）、SQL Injection Exploit（SQL 注入工具）等。可以采用的防范方法是为操作系统或应用程序打上相应缺陷的 Patch（补丁），恶意软件就无法利用 Exploit 对目标系统进行远程控制了。

2）Trojan Horse（特洛伊木马）

Trojan Horse 简称为 Trojan。它是伪装成合法程序以欺骗用户执行的一类恶意软件。Trojan 入侵目标系统的具体过程为：Trojan 首先通过网络或各种存储介质传播到用户端执行后，释放出其携带的 Backdoor（后门）以实现对目标主机的远程控制，在必要时还可释放出其携带的 Exploit 以提升用户权限。

除通过可执行文件复制、执行来传播的 Trojan 外，更具欺骗性的 Trojan 类型有邮件附件 Trojan、网页恶意代码 Trojan、宏病毒 Trojan 等，这几种 Trojan 往往在用户毫不知情的情况下就被执行了，危害非常大。

3）Worm（蠕虫）

它是具有自我繁殖能力，无须用户干预便可自动在网络环境中传播的一类恶意软件。Worm 利用目标系统的 Weak Password（弱密码）或目标系统中存在的缺陷获得对目标系统的远程控制权，并搜集目标系统内的相关信息，将 Worm 自身传染至与目标系统有网络联系的其他系统。

Worm 自身的存在有两种形式：可执行文件的形式和内存中进程/线程的形式。当以进程/线程的形式存在时，传播过程中的 Worm 在目标系统内不涉及文件操作，具有更强的隐蔽性。例如 Code Red Worm，其 Payload 部分的功能是对白宫 Web 服务器进行 DoS 攻击。

4）Bot（蠕虫）

Bot 的概念与 Worm 相近。若某 Worm 的 Payload 部分包含一个 Backdoor，则称该 Worm 为 Bot，或者说可远程控制的 Worm 即为 Bot。

Bot 比 Worm 更进一步，Bot 在传播出去之后依然可以对已入侵的主机进行控制和调整。由此可见 Bot 的危害性比 Worm 更大。

被 Bot 成功入侵后的受控主机即为 Zombie（傀儡主机），一组 Zombie 称为 Botnet。当前 Internet 上的主要安全威胁之一的 DDoS（Distributed Denial of Service，分布式拒绝服务）攻击就是通过向 Botnet 发出针对特定目标的 DoS 攻击命令来完成的。

5）Virus（病毒）

它是依附于宿主文件，在宿主文件被执行的条件下跟随宿主文件四处传播并完成特定业务功能的一类恶意软件。

第一类恶意软件对例如表 11-1 所示。

表 11-1 第一类恶意软件对比

病毒种类	Exploit	Trojan	Worm	Bot	Virus
获取目标系统控制权的能力	√	√	√	√	√
能否包含 Payload	×	√	√	√	√
感染目标系统模式	主动	被动	主动	主动	被动

注意：被动感染目标系统模式下，恶意软件需要目标系统用户执行外来程序才能获取对目标系统的远程控制权；主动感染目标系统模式下，恶意软件自行获取对目标系统的远程控制权，无须目标系统用户的任何动作。

2. 维持远程控制权类

获取对目标系统的远程控制权后，在目标系统中运行此类恶意软件以维持对目标系统的远程控制权。此类恶意软件用于完整入侵过程中的第二阶段。

1）Backdoor（后门）

它是一类运行在目标系统中为攻击者提供对目标系统未经授权的远程控制服务的恶意软件。Backdoor 必须在目标系统上运行才能提供相应的服务，因此必须先使用第一类恶意软件以获得在目标系统上执行程序的权限。第一类恶意软件是 Backdoor 发挥作用的前提和基础。

2）Rootkit

其概念起源于 UNIX/Linux 操作系统，最初是指 UNIX/Linux 系统中一组用于获取并维持 Root 权限的工具集。发展至今日，被广为接受的 Rootkit 概念是指用于帮助入侵者在获取目标主机管理员权限后，尽可能长久地维持这种管理员权限的工具。在当前的 Rootkit 概念中，获取管理员权限的过程不由 Rootkit 来完成，即 Rootkit 的使用基于已经获得了管理员权限的假设。

由 Backdoor 的概念可知，Backdoor 仅提供了一条非授权访问、控制目标系统的通道，但并不涉及对这条通道的保护，因此这条通道很容易被目标系统上的管理员或网络安全设备察觉或检测到。Rootkit 的作用是要尽可能长久地维持对目标系统的远程控制，故其基本任务就是要隐藏 Backdoor 所提供的通道，尽可能使目标系统上的管理员或安全设备不能察觉、检测到该通道的存在。

在实际应用中，Rootkit 通常直接包含了 Backdoor 的功能。因此可将 Rootkit 理解为带隐藏功能的 Backdoor。

第二类恶意软件对例如表 11-2 所示。

表 11-2 第二类恶意软件对比

恶意软件种类	Backdoor	Rootkit
提供非授权的远程访问控制通道	√	√
隐藏自身的能力	×	√
隐藏远程访问控制通道的能力	×	√
隐藏目标系统中运行的其他恶意软件	×	√

3. 完成特定业务逻辑类

第三类恶意软件用于完成入侵目标系统最终所要进行的操作，例如窃取情报、破坏系统、发动攻击、中转数据等。第一类和第二类恶意软件的作用在于为第三类恶意软件提供一个安全、便捷的运行平台。

1）Spyware（间谍软件）

它是典型的第三类恶意软件，用于从目标系统中收集各种情报、信息，例如商业、军事情报、用户信用卡号、各种网站/邮箱用户名和密码等信息。收集到这些信息后，Spyware 通过网络将其发送给入侵者。在 Rootkit 的保护下，Spyware 本身以及 Spyware 所产生的网络通信都被隐藏，使得 Spyware 在目标系统中安全地存活下来。

Spyware 在目标系统中的运行途径主要有 3 种：

（1）将 Spyware 放在 Worm/Bot/Virus/Trojan 的 Payload 中，在 Worm /Bot/Virus/Trojan 执行时被释放出来并执行。

（2）利用 Exploit 获取对目标系统远程控制权后，将 Spyware 通过网络传输到目标主机并执行。

（3）在目标系统上安装并运行 Rootkit/Backdoor 后，通过其提供的远程控制服务将 Spyware 传输到目标主机并执行。

2）Spamware（垃圾信息发送软件）

为了避免被追查，Spam（非期望的垃圾信息，例如垃圾邮件）的发送者通常不会直接使用自己的主机发送垃圾信息。为了加大垃圾信息的发送范围，仅仅用一台主机发送垃圾信息是不够的。Spamware 就是运行在大量被入侵主机中用于发送垃圾信息的恶意软件。Spamware 在目标系统中的运行途径与 Spyware 类似。

3）Adware（垃圾广告软件）

它是运行在被入侵主机中，用于以各种方式显示垃圾广告的恶意软件。Adware 在目标系统中的运行途径与 Spyware 类似。

4）其他第三类恶意软件

其种类很多，根据具体应用需求不同而不同，例如对目标系统进行攻击/破坏的恶意软件有多种不同的类型，但目前尚无统一规范的命名。

11.2　特洛伊木马

11.2.1　特洛伊木马介绍

古希腊传说中，特洛伊王子帕里斯访问希腊，诱走了王后海伦，希腊人因此远征特洛伊。围攻到第 10 年，希腊将领奥德修斯献了一计，把一批勇士埋伏在一匹巨大的木马腹内，放在城外后，佯作退兵。特洛伊人以为敌兵已退，就把木马作为战利品搬入城中。到了夜间，埋伏在木马中的勇士跳出来，打开了城门，希腊将士一拥而入攻下了城池。后世称这只大木马为特洛伊木马。

网络社会中的特洛伊木马并没有传说中的那样庞大，它们是一段精心编写的程序。与传说中的木马一样，它们会在用户毫不知情的情况下悄悄地进入用户的计算机，进而反客为

主,窃取机密数据,甚至控制系统。

特洛伊木马是指隐藏在正常程序中的一段具有特殊功能的恶意代码,是具备破坏和删除文件、发送密码、记录键盘和攻击 DoS 等特殊功能的后门程序。世界上第一个计算机木马是出现在 1986 年的 PC-Write 木马,它伪装成共享软件 PC-Write 的 2.72 版本(事实上,编写 PC-Write 的 Quicksoft 公司从未发行过 2.72 版本),一旦用户信以为真运行该木马程序,结果就是硬盘被格式化。

早期的木马和病毒的主要区别是病毒具有自传播性,而木马则不具备这一点。这主要是由于病毒的创造者当时主要是为了炫耀自己天才的创意,编制出可以自我复制传播并做出不可思议的效果(花屏、死机,甚至仅仅是一些有趣的画面),仿佛具有生命的计算机病毒,以极大地满足他们的成就感。与此相反,木马的创造者大多是有目的地破坏与损害特定软件的名誉,其中安插系统后门和盗取用户资料是典型的使用方式。

在 Internet 高度发达的今天,木马和病毒的区别正在逐渐变淡并消失。木马为了入侵并控制更多的计算机,糅合了病毒的编写方式,以获取更多的信息。因此,现在木马常被称为木马病毒。

木马程序一般都包括客户端和服务器端两个程序,其客户端是攻击者远程控制植入木马的机器,服务器端程序就是木马程序,如图 11-1 所示。攻击者要通过木马攻击受害者的系统,第一步是要把木马的服务器端程序植入受害者的计算机里。

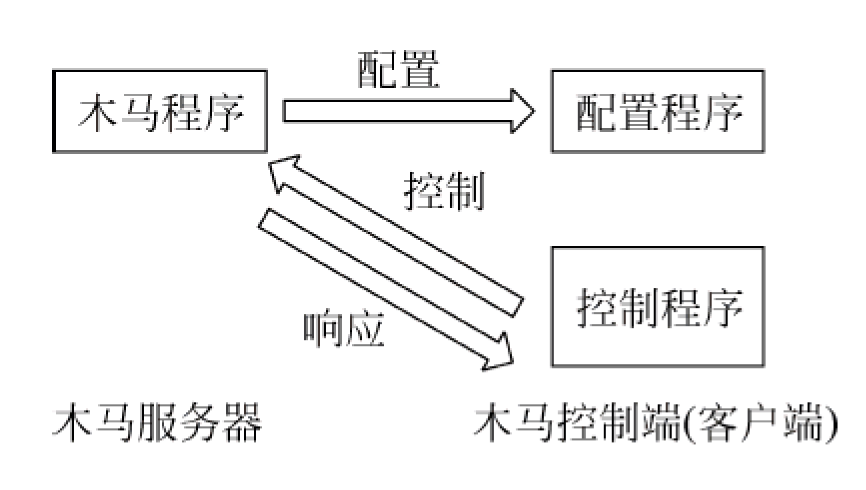


图 11-1 木马程序结构示意图

目前木马入侵的主要途径还是先通过一定的方法把木马执行文件植入受害者的计算机系统,可利用的途径有邮件附件、下载软件等。然后通过一定的提示故意误导被攻击者打开执行文件,例如故意谎称这个木马执行文件是来自朋友的贺卡。

木马也可通过 Script、ActiveX 及 Asp. CGI 交互脚本的方式植入。曾经出现过一个利用微软 Scripts 脚本漏洞对浏览者硬盘进行格式化的 HTML 页面植入方式。

11.2.2 特洛伊木马运行方式

作为一个优秀的木马,自启动功能是必不可少的,这样可以保证木马不会因为用户的一次关机操作而彻底失去作用。正因为该项技术如此重要,很多编程人员都在不停地研究和探索新的自启动技术,并且时常有新的发现。一个典型的例子就是把木马加入到用户经常执行的程序(例如 explorer. exe)中,用户执行该程序时,木马自动发生作用。当然,更加普遍的方法是修改 Windows 系统文件和注册表。

1. 自启动激活木马

防范自启动木马的条件,大致有以下 6 个方面:

(1) 注册表。打开 HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\Current Version 下的 5 个以 Run 和 RunServices 为主键的文件,在其中寻找可能是启动木马的键值。

(2) WIN. INI。C:\Windows 目录下有一个配置文件 win. ini,用文本方式打开,在[windows]

字段中有启动命令 load=和 run=，在一般情况下是空白的，如果有启动程序，可能是木马。

(3) SYSTEM. INI。C:\Windows 目录下有个配置文件 system. ini，用文本方式打开，在[386Enh]、[mci]、[drivers32]字段中有命令行，在其中寻找木马的启动命令。

(4) Autoexec. bat 和 Config. sys。在 C 盘根目录下的这两个文件也可以启动木马。但这种加载方式一般都需要控制端用户与服务器端建立连接后，将已添加木马启动命令的同名文件上传到服务器端覆盖这两个文件才行。

(5) *. INI。即应用程序的启动配置文件，控制端利用这些文件能启动程序的特点，将制作好的带有木马启动命令的同名文件上传到服务器端覆盖这些同名文件，这样就可以达到启动木马的目的了。

(6) 启动菜单：在"开始"→"程序"→"启动"选项下也可能有木马的触发条件。

2. 触发式激活木马

(1) 注册表。打开 HKEY_CLASSES_ROOT 文件类型\shellopencommand 主键，查看其键值。例如国产木马"冰河"，激活方法就是将注册表 HKEY_CLASSES_ROOT xtfileshellopencommand 下的键值"C:WINDOWS NOTEPAD. EXE"修改为"C:WINDOWSSYSTEMSYXXXPLR. EXE"，这时双击一个 txt 文件后，原本应用 Notepad 打开文件却变成启动木马程序了。

(2) 捆绑文件。实现这种触发条件首先要控制端和服务器端已通过木马建立连接，然后控制端用户用工具软件将木马文件和某一应用程序捆绑在一起，上传到服务器端覆盖原文件。这样即使木马被删除了，只要运行捆绑了木马的应用程序，木马又会被安装上来。

(3) 自动播放。自动播放原本是用于光盘的，其本意是插入一个电影光盘到光驱时，系统会自动播放里面的内容。后来有人在 U 盘或硬盘的分区创建 Autorun. inf 文件，并在 Open 中指定木马程序。这样，当用户打开硬盘分区或 U 盘时，就会触发木马程序的运行。

木马被激活后，进入内存，开启事先定义的木马端口准备与控制端建立连接。这时，服务端用户可以在 MS-DOS 方式下，输入"netstat-an"查看端口状态。个人计算机在脱机状态下一般是不会有端口开放的，如果有端口开放，就要注意是否感染木马了。

11.2.3　木马的隐藏性

特洛伊木马的危害在于它对计算机系统具有强大的控制和破坏能力。木马之所以会造成很大损失，其根本原因就是其隐蔽性非常强。木马的隐藏性也是病毒的最大特点。下面对木马病毒的隐藏方式进行分析。

1. 在任务栏里隐藏

这是最基本的隐藏方式。如果在 Windows 的任务栏里出现一个莫名其妙的图标，用户一眼就会发现异常。

2. 在任务管理器里隐藏

可以通过运行任务管理器来查看正在运行的进程，但是木马程序往往不会这么轻易被发现。

3. 端口

一台机器有 65 536 个端口。经分析，大多数木马使用的端口在 1024 以上，并且数字呈

越来越大的趋势。当然也有占用1024以下端口的木马,但是由于这些端口是常用端口,占用这些端口可能会造成系统不正常,木马会很容易暴露。

4. 隐藏通信

隐藏通信也是木马经常采用的手段之一。任何木马运行后都要和攻击者进行通信连接。可以通过即时连接,例如攻击者通过客户端直接接入被植入木马的主机。也可通过间接通信,例如木马通过电子邮件的方式将被入侵主机的敏感信息发送给攻击者。

5. 隐藏加载方式

木马加载的方式可以说千奇百怪,无奇不有。随着网站互动化的不断进步,越来越多的载体可以成为木马的传播介质,例如JavaScript、VBScript、ActiveX、XLM等。几乎网络的每一个新功能都会导致木马的快速进化。

6. 最新隐身技术

在研究了其他软件的长处之后,攻击者发现Windows下的中文汉化软件采用的陷阱技术非常适合木马的使用。

这是一种更新、更隐蔽的方法。通过修改虚拟设备驱动程序(VxD)或修改动态链接库(DLL)来加载木马。它基本上摆脱了原有的木马模式(监听端口),采用替代系统功能的方法(改写VxD或DLL文件),木马会将修改后的DLL替换系统已知的DLL,并对所有的函数调用进行过滤。对于常用的调用,使用函数转发器直接转发给被替换的系统DLL,一旦发现木马控制端的请求就激活自身。这样做的好处是没有增加新的文件,不需要打开新的端口,没有新的进程,使用常规的方法监测不到木马。当木马的控制端向被控制端发出特定的信息后,隐藏的程序才会立即开始运作。

11.2.4 常见特洛伊木马介绍

1. 破坏型

唯一的功能就是破坏并删除文件,可以自动地删除计算机上的DLL、INI、EXE文件。

2. 密码发送型

可以找到隐藏的密码并把它们发送到指定的信箱。有些用户喜欢把自己的各种密码以文件的形式存放在计算机中,认为这样方便;还有用户喜欢用Windows提供的密码记忆功能,这样就可以不必每次都输入密码。许多黑客软件可以寻找到这些文件,把它们发送到黑客手中。也有些黑客软件长期潜伏,记录用户的键盘操作,从中寻找有用的密码。

3. 远程访问型

如果用户运行了服务器端程序,攻击者只需知道服务器端的IP地址就可以实现远程控制。

4. 键盘记录木马

这种特洛伊木马非常简单。它们只做一件事情,就是记录受害者的键盘敲击行为并且在LOG文件里查找密码。这种特洛伊木马随着Windows的启动而启动,它们具有在线和离线记录两种选项,分别记录用户在线和离线状态下敲击键盘时的按键情况。当然,对于这种类型的木马,邮件发送功能也是必不可少的。

5. DoS 攻击木马

当前，用作 DoS 攻击的木马越来越流行。当攻击者入侵了一台机器，植入 DoS 攻击木马，这台机器就成为了攻击者发动 DoS 攻击的得力助手。攻击者控制的傀儡机器数量越多，发动 DoS 攻击取得成功的概率就越大。因此，这种木马的危害不是体现在被感染的机器上，而是体现在攻击者可以利用它来攻击一台又一台计算机，给网络造成巨大的危害。

6. 代理木马

黑客在入侵的同时掩盖自己的足迹，防止身份被发现是非常重要的。因此，给被控制的傀儡机器种上代理木马，通过代理木马，攻击者可以在匿名的情况下使用 ICQ、Telnet、IRC 等程序，从而隐蔽自己的踪迹。

7. FTP 木马

这种木马可能是最简单最古老的木马了，它的唯一功能就是打开 21 端口，等待用户连接。

11.2.5 防范木马的安全建议

一些常见的木马，例如 SUB7、BO2000、冰河等，由于它们都是采用打开 TCP 端口监听和写入注册表启动等方式运行，使用木马克星之类的软件可以检测到这些木马。下面介绍几种在使用计算机时应当注意的非正常情况。

(1) 浏览器突然自动打开，并且进入某个网站。

(2) 操作计算机时，突然弹出一个警告框或询问框。

(3) Windows 系统配置总是自动莫名其妙地被更改。例如屏保显示的文字、时间和日期、声音大小、鼠标灵敏度以及 CD-ROM 的自动运行配置。

(4) 硬盘无缘由地读盘，软驱灯经常自己亮起，网络连接及鼠标、屏幕出现异常现象。

最简单的方法就是使用 netstat-a 命令查看。用户可以通过这个命令发现所有的网络连接，如果此时有攻击者通过木马连接用户机器，就可以通过这些信息发现异常。使用端口扫描的方法也可以发现一些简单的木马。

当然，没有上述的现象并不代表用户机器就绝对安全。有些攻击者攻击受害者的机器不过是想寻找一个跳板。对于那些隐藏得很深的木马，用户的检查工作将变得异常艰难，并且需要对入侵和木马有较强的敏感度。

如果用户怀疑自己的机器已经被木马入侵，就需要马上采取以下措施：所有的账户和密码都要马上更改；删掉所有硬盘上原来没有的东西；检查硬盘上是否有病毒存在。然后再进行更深一步的处理。

11.3 计算机病毒

11.3.1 什么是计算机病毒

1. 计算机病毒的概念

通常所说的计算机病毒，在《中华人民共和国计算机系统安全保护条例》中有明确定义：

病毒是指“编制者在计算机程序中插入的破坏计算机功能或者破坏数据，影响计算机使用并且能够自我复制的一组计算机指令或者程序代码”。这些程序之所以称为病毒，主要是由于它们与生物医学上的病毒有着很多相似点。例如，它们都具有寄生性、传染性和破坏性，有些恶意代码会像生物病毒隐藏和寄生在其他生物细胞中那样寄生在计算机用户的正常文件中，伺机发作并大量地复制病毒体，感染本机的其他文件和网络中的计算机。绝大多数的恶意代码都会对人类社会生活造成不利的影响，造成的经济损失数以亿计。

与医学上的病毒不同，计算机病毒不是天然存在的，是某些程序编写者利用计算机软件和硬件所固有的脆弱性编制的一组指令集或程序代码。它能通过某种途径潜伏在计算机的存储介质(或程序)里，达到某种条件时即被激活，通过修改其他程序的方法将自己的精确复制或者可能演化的形式放入其他程序中，从而感染其他程序，对计算机资源进行破坏。

从广义上来说，凡能够引起计算机故障，破坏计算机数据的程序统称为计算机病毒。依据此定义，诸如逻辑炸弹、蠕虫等均可称为计算机病毒。计算机病毒与人们平时所使用的各种软件程序从本质上看并没有什么区别，但正常的程序或软件是用来帮助人们解决某些问题的，而病毒是专门用来搞破坏的，即病毒程序是一种有害的程序。

2. 计算机病毒的结构

计算机病毒虽然长短不同、大小各异，但它们的主要结构一般是类似的，都包含四大模块：引导模块、触发模块、感染模块、表现(或破坏)模块，如图 11-2 所示。

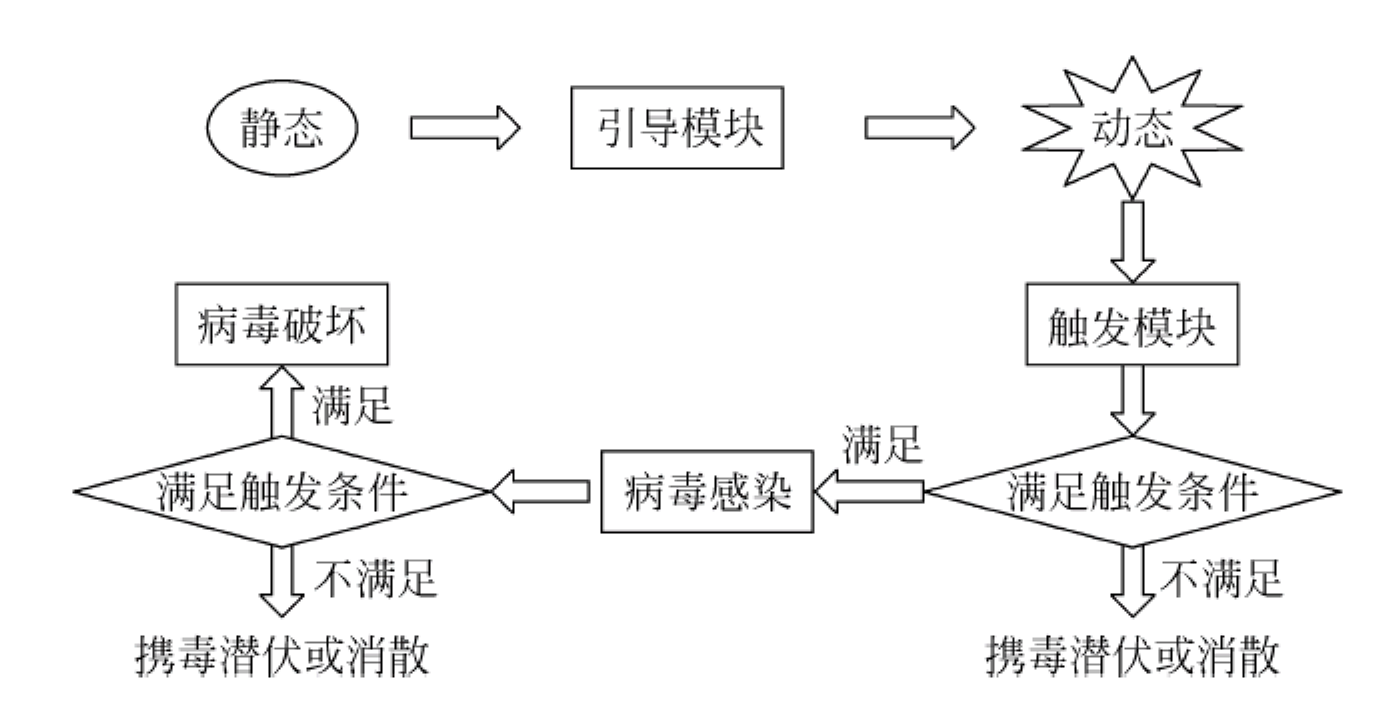


图 11-2 计算机病毒主要结构示意图

(1) 引导部分的作用是将病毒主体加载到传染目标上去，为传染部分做准备，例如驻留内存、修改中断、修改注册表等操作。

(2) 传染部分的作用是将病毒代码复制到传染目标上去。不同类型的病毒的传染方式、传染条件以及传播所借助的媒体各不同。

(3) 表现部分是病毒间差异最大的部分，前两部分是为这部分服务的。大部分的病毒都要满足一定条件才会触发其表现部分，例如以计算机时钟作为触发条件或用键盘输入特定字符来触发。

(4) 后两个部分各包含一段触发条件验证代码，当各段验证代码分别验证出传染和表现(或破坏)的触发条件满足时，病毒就会进行传染和表现(或破坏)。必须指出的是，不是任何病毒都包含这 3 个部分。

3. 计算机病毒的命名

给病毒命名是病毒研究和反病毒技术的一部分，计算机用户通常知道的病毒名称主要是由各个反病毒产品厂家命名的。由于反病毒厂家很多，有时对同一种病毒命名为不同的名称，例如“SPY”病毒，KILL 系列将其命名为 SPY，瑞星则叫“3783”。病毒的命名并没有一个统一的规定，每个反病毒公司的命名规则都不太一样，但基本都是采用前、后缀法来进行命名的，可以是多个前缀、后缀组合，中间以小数点分隔，一般格式为〔前缀〕.〔病毒名〕.〔后缀〕。

1）病毒前缀

病毒前缀是指一个病毒的种类，常见的有 Script（代表脚本病毒）、Trojan（代表木马病毒）、Worm（代表蠕虫病毒）、Harm（代表破坏性程序）、Macro/WM/WM97/XM/XM97（代表宏病毒）、Win32/W32（代表系统病毒）。一般 DOS 类型的病毒是没有前缀的。

2）病毒名

病毒名是指病毒的名称。例如以前很有名的 CIH 病毒，它和它的一些变种都统一用 CIH。还有振荡波蠕虫病毒，它的病毒名是 Sasser。常用的病毒命名方法有如下几种：

（1）按病毒出现的地点命名，例如 chongqin_JES 表示其样本最先出现在重庆的某用户计算机中。

（2）按病毒中出现的人名或特征字体命名，例如 ZHANGFANG-1535。

（3）按病毒发作时的症状命名，例如毛毛虫、火炬。

（4）按病毒发作的时间命名，例如 NOVEMBER 9TH 病毒每逢 11 月 9 日发作。

（5）有些名称则包含病毒代码的长度，例如 PLXEL. ×××系列、KO. ×××等。

3）病毒后缀

病毒后缀是指一个病毒的变种特征，一般是采用英文中的 26 个字母来表示的，例如 Nimda 病毒命名为 Worm. Nimda. i。如果病毒的变种太多了，也可采用数字和字母混合的方法来表示病毒的变种。

11.3.2　计算机病毒的特征

计算机病毒赖以生存的基础是现代计算机均采用了冯·诺依曼的存储程序工作原理和操作系统的公开性和脆弱性，以及网络中的漏洞。程序和数据都存在计算机中，它们都可以被读、写、修改和复制，即程序可以在内存中繁殖。计算机病毒有许多特征，按照目前信息安全领域的普遍观点，计算机病毒的常见特征有非法性、隐蔽性、潜伏性、可触发性、可执行性、破坏性、传染性和针对性。

需要指出的是，单独根据以上的某一特征是不能判断某个程序是否是病毒的。以破坏性为例，DOS 操作系统的 Format 程序虽然能消除磁盘上的数据，造成对数据的破坏，但它显然不是病毒，因为它除了不具备病毒的传染性这个根本特征外，也不具有病毒的其他大部分特征。

1. 非法性

正常情况下，计算机用户调用执行一个合法程序时，把系统控制权交给这个程序，并为其分配相应的系统资源使之能运行，以达到用户的目的，程序执行的过程对用户来说是透明

和可知的，因此这种程序是“合法”的。

然而计算机病毒是非法程序，计算机用户不会明知是病毒程序而故意去执行它。由于病毒具有正常程序的一切特征，它会将自己隐藏在合法的程序或数据中，当用户执行正常合法程序时，病毒伺机窃取到系统的控制权，得以抢先运行，而用户还认为在执行正常程序。由此可见，病毒的行为都是在未获得用户的允许下悄悄进行的，病毒所进行的操作绝大多数都是违背用户意愿和利益的。因此，计算机病毒具有非法性。

2. 隐蔽性

病毒一般是具有很高编程技巧、短小精悍的程序，通常附着在正常程序中或磁盘较隐蔽的地方，也有个别病毒以隐含文件形式出现，目的都是不让用户发现它的存在。如果不进行代码分析，病毒程序与正常程序是不容易区别开来的。正是由于具有隐蔽性，计算机病毒才能在用户没有察觉的情况下快速扩散到网络上数百万台计算机中。

不同类型病毒的隐藏方式也是多种多样的。有些病毒隐藏在磁盘上标为坏簇的扇区以及一些空闲概率较大的扇区中，也有个别的病毒以隐含文件的形式出现。

3. 潜伏性

计算机病毒潜伏性的第一种表现是不使用专用检测程序是检查不出病毒程序的，它可以在磁盘里潜伏几天甚至几年，一旦时机成熟得到运行机会，就会四处繁殖与扩散。潜伏性的第二种表现是其内部往往有一种触发机制，不满足触发条件时，计算机病毒除了传染外不做其他破坏。触发条件一旦得到满足，有的会在屏幕上显示信息或特殊标识，有的则执行破坏系统的操作，例如格式化磁盘、删除磁盘文件等。著名的“黑色星期五”的触发条件就是日期是 13 号的星期五。

4. 可触发性

病毒因某个事件或数值的出现诱使其实施感染或进行攻击的特性称为可触发性。为了隐蔽自己，病毒必须潜伏在机器内少做动作。如果完全不动，病毒既不能感染也不能进行破坏，就失去了杀伤力。病毒既要隐蔽又要维持杀伤力，就必须具有可触发性。病毒的触发机制是用来控制感染和破坏动作的频率的。病毒具有预定的触发条件，这些条件可以是时间、日期、文件类型或某些特定数据等。

5. 可执行性

计算机病毒与其他合法程序一样，是一段可执行程序，但它不是一个完整的程序，而是寄生在其他可执行程序上。在病毒运行时，病毒程序与合法程序争夺系统的控制权。

只有当计算机病毒在计算机内得以运行时，才具有传染性和破坏性等活性。因此，计算机病毒实施危害的关键是获得计算机 CPU 的控制权。

6. 破坏性

所有的计算机病毒都是一种可执行程序，因此对系统来说，所有的计算机病毒都存在一个共同的危害，即降低计算机系统的工作效率，占用系统资源，其具体情况取决于入侵系统的病毒程序。同时，计算机病毒的破坏性主要取决于计算机病毒设计者的目的，如果病毒设计者的目的在于彻底破坏系统的正常运行的话，那么这种病毒对于计算机系统进行攻击造成的后果是难以设想的，它可以毁掉系统的部分数据，也可以破坏全部数据并使之无法恢

复。某些本没有多大破坏作用的病毒交叉感染，也会导致系统崩溃等重大恶果。

7. 传染性

传染性是生物病毒的基本特征，也是判断一段程序代码是否是计算机病毒的依据。

计算机病毒一旦进入计算机并得以执行，就会搜寻其他符合其传染条件的程序或存储介质，确定目标后再将自身代码插入其中，达到自我繁殖的目的。只要一台计算机染毒，若不及时处理，病毒就会在这台机器上迅速扩散。其中的大量文件(一般是可执行文件)会被感染，而被感染的文件又成了新的传染源，再与其他机器进行数据交换或通过网络接触，病毒就会得以继续传染。

8. 针对性

计算机病毒一般都是针对特定的操作系统，例如微软的 XP 和 Windows 7，也有针对特定应用程序的计算机病毒。这种针对性有两个特点：如果对方是它针对的操作系统类型，完全获得对方的管理权限后就可以肆意妄为；反之这种病毒就会失效。

11.3.3 计算机病毒的分类

自第一个计算机病毒问世以来，计算机病毒的数量逐年增加。根据每年从计算机用户反馈的有关病毒的信息分析，从 20 世纪 90 年代初的每月几种达到了现在的每天 200 种以上。

由于计算机病毒及其所处环境的复杂性，以某种方式遵循单一标准为病毒分类的方法无法达到对病毒的准确认识，也不利于对病毒的分析与防治，下面从多个角度对计算机病毒进行详细分类。

1. 根据寄生的数据存储方式划分

计算机病毒根据寄生的数据存储方式可划分为 3 种类型：引导型病毒、文件型病毒和混合型病毒。

1）引导型病毒

引导型病毒是指寄生在磁盘引导区或主引导区的计算机病毒。这种病毒利用操作系统引导模块在某个固定位置并且控制权的转交方式以物理位置为依据，系统不对主引导区的内容正确与否进行判别的缺点，在引导系统的过程中将真正的引导区内容转移，待病毒程序执行后，将控制权交给真正的引导区内容，从而完成侵入系统、驻留内存、监视系统运行、待机传染和破坏行为。按照引导型病毒在硬盘上的寄生位置又可细分为主引导记录病毒和分区引导记录病毒。主引导记录病毒感染硬盘的主引导区，例如大麻病毒、2708 病毒、火炬病毒等；分区引导记录病毒感染硬盘的活动分区引导记录，例如小球病毒、Girl 病毒等。

2）文件型病毒

文件型病毒主要感染计算机中的可执行文件(.exe)和命令文件(.com)。文件型病毒是对计算机的源文件进行修改，使其成为新的带毒文件。一旦计算机运行该文件就会被感染，从而达到传播的目的。

文件型病毒分两类：一种是将病毒加在 com 文件的前部；另一种是将病毒加在文件尾部。

将所有通过操作系统的文件系统进行感染的病毒都称作文件型病毒，因此这是一类数

目非常巨大的病毒。理论上可以制造这样一个病毒，该病毒基本上可以感染所有操作系统的可执行文件。目前已经存在这样的文件病毒，可以感染所有标准的DOS可执行文件，包括批处理文件、DOS下的可加载驱动程序(.sys)文件以及普通的com/exe可执行文件。当然还有感染所有Windows操作系统可执行文件的病毒，可感染文件的种类包括Windows 3.x版本、Windows 9x版本、Windows NT和Windows 2000版本下的可执行文件，后缀名是exe、dll或vxd、sys。

除此之外，还有一些病毒可以感染高级语言程序的源代码、开发库和编译过程所生成的中间文件。

3）混合型病毒

混合型病毒是指具有引导型病毒和文件型病毒两种寄生方式的计算机病毒，又称综合性或复合性病毒，所以它的破坏性更大，传染的机会也更多，杀灭也更困难。这种病毒扩大了病毒程序的传染途径，它既感染磁盘的引导记录，又感染可执行文件。当染有此种病毒的磁盘用于引导系统或调用执行染毒文件时，病毒都会被激活。因此在检测、清除复合型病毒时，必须全面彻底地根治。如果只发现该病毒的一个特性，把它只当作引导型或文件型病毒进行清除，虽然好像是清除了病毒，但还留有隐患，这种经过消毒后的“洁净”系统更赋有攻击性。这类病毒有Flip病毒、新世纪病毒、One-half病毒等。

2. 根据感染类型划分

保守地讲，计算机系统文件类型在300种以上。若按扩展名来说，目前能被病毒感染的文件包括exe、com、dll、sys、vxd、drv、bin、ovl、386、htm、fon、doc、dot、xls、xlt、vbs、vbe、js、jse、css、wsh、sct、hta、htt、asp、zip、arj、cab、rar、zoo、arc、lzh、pkzip、gzip、pkpak、ace等文件。虽然被感染的表现形式不一样，但从本质上讲，病毒都是感染文件的程序指令代码部分。

3. 根据病毒攻击的计算机类型划分

1）微型计算机病毒

微型计算机病毒是世界上传染最为广泛的一种病毒。

2）小型机计算机病毒

小型机计算机病毒的应用范围是极为广泛的，它既可以作为网络的一个节点机，也可以作为小的计算机网络的主机。起初人们认为计算机病毒只有在微型计算机上才能发生，而小型机则不会受到病毒的侵扰。但是自从1988年11月Internet受到蠕虫病毒的攻击后，人们认识到小型机同样不能免遭计算机病毒的攻击。

3）工作站病毒

近几年来，计算机工作站技术有了较大的进展，应用范围也有了较快发展。不难想象，攻击计算机工作站的病毒的出现也是对信息系统的一大威胁。

4. 根据病毒的破坏程度划分

1）良性病毒

良性病毒是指那些只是为了表现自身，并不彻底破坏系统和数据，但会大量占用CPU时间，增加系统开销，降低系统工作效率的一类计算机病毒。这种病毒多数是恶作剧者的产物，例如小球病毒、1575/1591病毒、救护车病毒、Rose病毒等。

有些用户对这类计算机病毒的传染不以为然。其实良性和恶性都是相对而言的。例

如，原本只有 10KB 的文件变成约 90KB，原因就是被几种病毒反复感染了数十次。这不仅消耗大量宝贵的磁盘存储空间，整个计算机系统也由于多种病毒寄生于其中而无法正常工作。因此不能轻视所谓良性病毒对计算机系统造成的损害。

2）恶性病毒

恶性病毒是指那些一旦发作就会破坏系统或数据，造成计算机系统瘫痪的一类计算机病毒。这类病毒有黑色星期五病毒、火炬病毒、米开朗·基罗病毒等。这类病毒危害性极大，发作后可能给用户造成不可挽回的损失。

5. 根据传播途径分类

1）单机病毒

单机病毒的载体是磁盘、光盘和 U 盘等可移动存储介质。常见的是病毒从软盘、U 盘、光盘等传入硬盘，感染操作系统及已安装的软件或程序，然后再通过存储介质的转移又传染其他系统。

2）网络病毒

网络病毒的传播媒介不再是移动式载体，而是网络数据通道。这种病毒的传染能力更强，破坏力更大。

6. 根据激发机制划分

1）实时发作型病毒

实时发作型病毒会在病毒程序或带毒文件被执行时，无条件限制而立即进行相应的感染和破坏。

2）间歇发作型病毒

有的病毒会潜伏一段时间，直到在它所设置的日期才发作。有的病毒发作时会在屏幕上显示一些带有宣示或警告意味的信息。病毒发作后可能会摧毁分区表，导致无法启动，或者直接格式化硬盘。

7. 根据病毒自身变化性划分

1）原形病毒

原形病毒在病毒传染和破坏过程中自身(病毒程序)不发生变化。

2）变型性病毒(或称为幽灵病毒)

变型性病毒会在每次进行感染的时候针对其新宿主的状况编写新的病毒代码，然后才进行感染。因此这种病毒没有固定的病毒代码。以扫描病毒代码的方式来检测病毒的查毒软件遇上这种病毒将毫无作用。但反病毒软件随着病毒技术的发展也在发展，现已具有对付这种病毒的有效手段。

8. 根据与被感染对象的关系划分

1）寄生病毒

这类病毒在感染宿主文件或引导区后会把自身代码与宿主程序融合在一起，通常感染病毒后的文件会变大。

2）伴随型病毒

这类病毒并不改变文件本身，它们根据算法产生 exe 文件的伴随体，具有同样的名字和不同的扩展名(com)，例如 xcopy.exe 的伴随体是 xcopy.com。病毒把自身写入 com 文件

而不改变 exe 文件，当 DOS 加载文件时，伴随体优先被执行，然后再由伴随体加载执行原来的 exe 文件。

3）独立型病毒

这类病毒通常不把自身代码加入到宿主文件，而是通过修改某一系统文件，使操作系统在某种条件下自动执行独立存储于磁盘中的病毒程序。

11.3.4 几种典型计算机病毒的分析

1. CIH 病毒

CIH 病毒是一种文件型病毒，又称 Win95. CIH、Win32. CIH、PE_CIH，是感染 Windows 95/98 环境下 PE 格式文件的病毒。CIH 病毒一共有 5 个版本，目前最流行的是 V1.2 版。CIH 病毒的特点是没有改变宿主文件的大小，而是利用“空洞”，将病毒化整为零，拆分成若干块，插入到宿主文件中。CIH 病毒发作时直接向计算机主板 BIOS 芯片和硬盘写乱码，造成主机无法启动。CIH 病毒首开攻击计算机硬件之先河，也开启了 Windows 病毒的新纪元。

查找 CIH 病毒最简单的方法就是在资源管理器中搜索包含特征字符串 CIHv 的所有 .exe 文件。

1）VxD 技术

VxD(Virtual Device Drivers，虚拟设备驱动)是微软专门为 Windows 制定的设备驱动程序接口规范。很多应用软件都需要使用 VxD 机制来实现某些比较特殊的功能。由于 VxD 程序具有比其他类型应用程序更高的优先级，而且更靠近系统底层资源，反病毒软件要利用 VxD 才有可能全面、彻底地控制系统资源。

CIH 病毒利用 VxD 技术，通过 Windows 9x 的异常处理机制(Exception)进入系统 Ring0 级，在应用程序下故意产生一个异常，并修改 IDT 表(中断地址表)中的处理程序地址，使其指向病毒代码，再显式进入此异常(主要为直接调用 INT 3H 中断)，就可以申请系统共享内存，将病毒驻留。

2）CIH 病毒的驻留(初始化)

当运行感染了 CIH 病毒的 PE 文件时，由于该病毒修改了该程序的入口地址，从而先调入内存执行，其驻留主要过程为：

(1) 用 SIDT 指令取得 IDT Base Address(中断描述符表基地址)，然后把 IDT 的 INT 3H 的入口地址改为指向 CIH 自己的 INT 3H 程序入口部分。

(2) 执行 INT 3H 指令，进入 CIH 自身的 INT 3H 入口程序，这样 CIH 病毒就可以获得 Windows 最高级别的权限 Ring0。病毒在这段程序中首先检查调试寄存器 DR0 的值是否为 0，用以判断先前是否有 CIH 病毒已经驻留。

(3) 如果 DR0 的值不为 0，则表示 CIH 病毒程式已驻留，病毒程序恢复原先的 INT 3H 入口，然后正常退出 INT 3H，跳到(9)。

(4) 如果 DR0 值为 0，则 CIH 病毒将尝试进行驻留。

(5) 如果内存申请成功，则从被感染文件中将原先分成多块的病毒代码收集起来，组合后放到申请到的内存空间中。

(6) 再次调用 INT 3H 中断进入 CIH 病毒体的 INT 3H 入口程序，调用 INT 20H 来完

成调用一个 IFSMgr_InstallFileSystemApiHook 的子程序，在 Windows 内核中文件系统处理函数中挂接钩子，以截取文件调用的操作，这样一旦系统出现要求开启文件的调用，则 CIH 病毒的传染部分程序就会在第一时间截获此文件。

(7) 将同时获取的 Windows 操作系统默认的 IFSMgr_Ring0_FileIO(核心文件输入/输出)服务程序的入口地址保留在 DR0 寄存器中，以便 CIH 病毒调用。

(8) 恢复原先的 IDT 中断表中的 INT 3H 入口，退出 INT 3H。

(9) 根据病毒程序内隐藏的原文件的正常入口地址，跳到原文件正常入口，执行正常程序。

3) CIH 病毒的感染

CIH 病毒的传染部分实际上是病毒在驻留内存过程中调用 Windows 内核底层函数 IFSMgr_InstallFileSystemApiHook 挂接钩子时指针指示的那段程序，其感染过程为：

(1) 文件的截获。每当系统出现要求开启文件的调用时，驻留内存的 CIH 病毒就截获该文件。病毒调用 INT20 的 VxD call UniToBCSPath 系统功能调用取回该文件的名称与路径。

(2) exe 文件的判断。对该文件名进行分析，若文件扩展名不为.exe 或不满足 PE 格式、尚未感染等条件，则不传染，离开病毒程序，跳回到 Windows 内核的正常文件处理程序上。

4) CIH 病毒的寄生方法

CIH 病毒的寄生方法如图 11-3 所示。

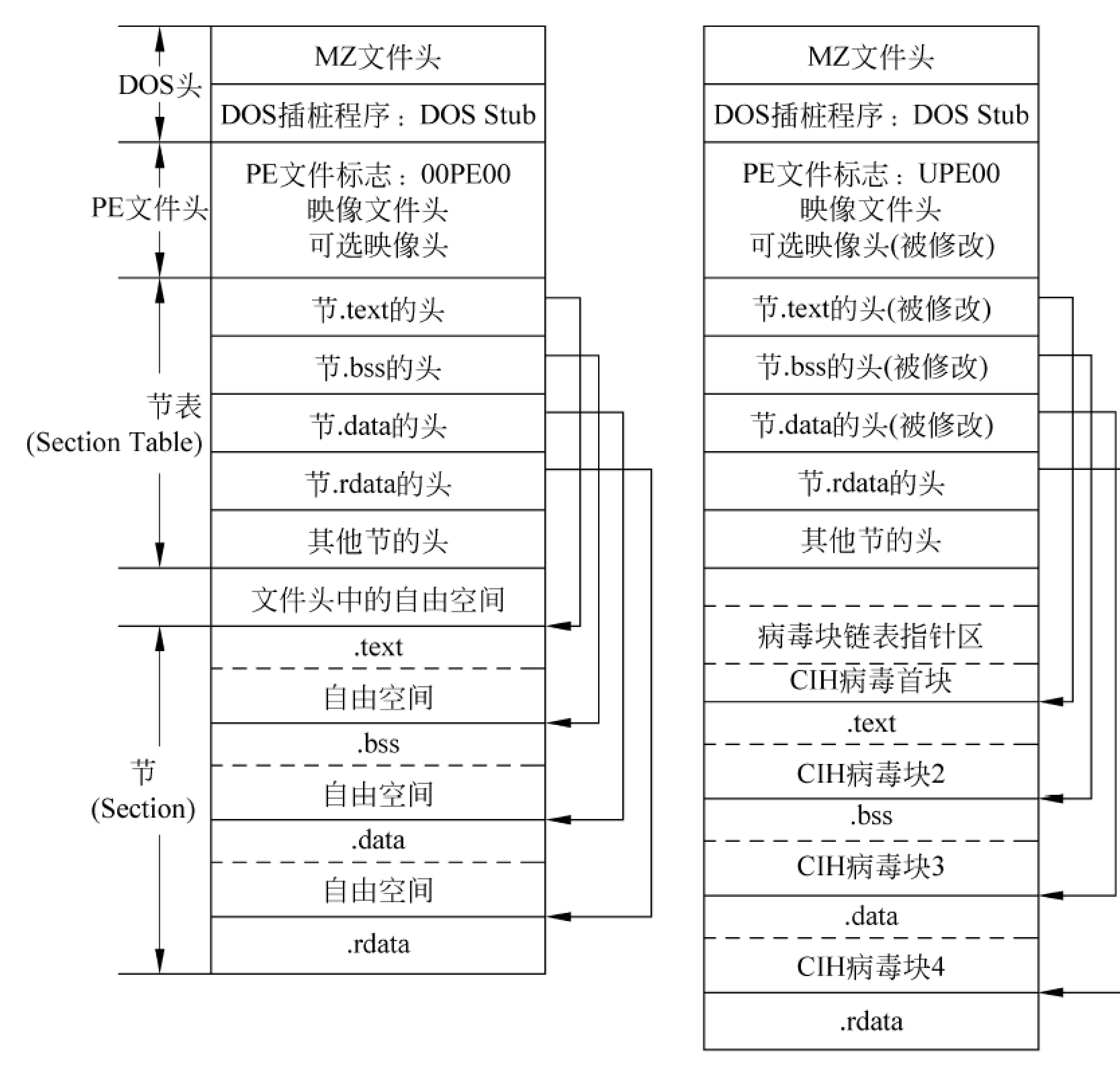


图 11-3　CIH 病毒的寄生方法

2. 爱虫病毒

爱虫(VBS. LoveLetter)别名 LoveLetter、LoveBug、VeryFunny,采用 VBScript 编制。它通过 Microsoft Outlook 将名称为 LOVE-LETTER-FOR-YOU. TXT. vbs 的邮件发送给用户地址簿里所有的地址,也可以产生 Script. ini 文件,将包括病毒的 LOVE-LETTER-FOR-YOU. HTM 文件通过 mIRC 蔓延到 Internet 聊天室中。病毒还有多个发作破坏模块,查找本地驱动器和网络驱动器,并在所有目录和子目录中搜索可以感染的目标,能够感染. vbs、. vbe、. js、. jse、. css、. wsh、. sct、. hta、. jpg、. jpeg、. wav、. txt、. gif、. doc、. htm、. html、. xls、. ini、. bat、. com、. mp3 和. mp2 等扩展文件,它用病毒代码覆盖文件原有的内容,并在文件名后面添加. vbs 的后缀名,从而取代宿主文件。其中,. mp2 和. mp3 文件被隐藏起来,实际上并未毁坏。

爱虫病毒一旦运行,会在系统中留下\windows\Win32DLL. vbs、\system\MSKernel. vbs 和\system\LOVE-LETTER-FOR-YOU. TXT. vbs 文件

1) 爱虫病毒各重要模块

(1) Main()。主模块,集成调用其他各个模块。

(2) regruns()。该模块主要用来修改注册表 Run 下面的启动项指向病毒文件和下载目录,并负责随机从给定的 4 个网址中下载 WIN_BUGSFIX. EXE 文件,并使启动项指向该文件。

(3) html()。生成 LOVE-LETTER-FOR-YOU. HTM 文件,该 HTM 文件执行后会执行其中的病毒代码,并在系统目录生成一个病毒副本 MSKernel32. vbs 文件。

(4) spreadtoemail()。将病毒文件作为附件发送给 Outlook 地址簿中的所有用户。

(5) listadriv()。搜索本地磁盘,并对磁盘文件进行感染。

2) 爱虫病毒的解毒步骤

(1) 将 wscript 文件"结束任务"。

(2) 执行 msconfig. exe 进入"启动"菜单,将所有的后缀为. vbs 的文件选择为禁用状态。

(3) 重新启动计算机。

(4) 查找一个 WIN-BUGFIX. exe 的文件并删除它,如果安装了 mIRC,则删除 script. ini 文件,删除含 LOVE-LETTER-FOR-YOU. TXT 附件的 E-mail。

(5) 打开注册表编辑器并且删除下列键值:

HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Run\MSKernel32;

HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\RunServer\Win32DLL;

HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Run\WIN_BUGSF IX。

11.3.5 计算机病毒的预防与清除

计算机病毒会导致大量的存储介质被感染、数据丢失和破坏,甚至整个计算机系统崩溃。计算机病毒严重地威胁着计算机信息系统的安全,如何有效地预防和控制计算机病毒

的产生、蔓延，清除入侵到计算机系统内的计算机病毒，是计算机工作人员研究的重大课题之一。

1. 病毒的预防

防止计算机病毒就像人类防止传染病一样，堵塞计算机病毒传播渠道是防止计算机病毒传染最有效的方法。堵塞病毒传播渠道的有效措施有以下几个方面：

（1）对公用软件和共享软件的使用要谨慎。

（2）对新添置的计算机系统要进行病毒检测。

（3）对所有不再需要写入数据的磁盘都应加写保护。

（4）系统数据经常做备份，检查、消毒后保存备用。

（5）一旦发现有计算机遭受病毒感染，应立即隔离，尽快杀毒。

（6）计算机网络上可执行文件的交换极易传播病毒，应限制网上可执行文件的交换。

2. 病毒的消除

由于病毒的破坏力越来越强，几乎所有的软、硬件故障都可能与病毒有关，因此当发现计算机有异常情况时，首先应该想到有可能是病毒在作怪。消除计算机病毒一般有两种方法：

（1）人工消除法。用工具软件对系统进行检测，消除计算机病毒。

（2）软件消除法。利用专门的防病毒软件，对计算机病毒进行检测和消除。

11.4　蠕虫病毒

11.4.1　什么是蠕虫病毒

蠕虫病毒是产生于 20 世纪 70 年代的较为古老的病毒。由于蠕虫病毒一开始便是根植于网络的，因此随着网络的发展，蠕虫病毒的生命力越来越强，其破坏力也越来越大。

早期的蠕虫不属于病毒，也不具备破坏性，它只是一种网络自动工具。

1982 年，Shock 和 Hupp 根据 *The Shockwave Rider* 一书中的概念提出了一种蠕虫（Worm）程序的思想。蠕虫程序可用作 Ethernet（以太网）网络设备的一种诊断工具，它能快速有效地检测网络。

1988 年 11 月 2 日，美国康乃尔大学学生罗伯特·莫里斯（Robert Morris）正是利用 UNIX 操作系统寄发电子邮件的公用程序中的一个缺陷，把他首创的人工生命蠕虫病毒放进 Internet 网络，闯下了弥天大祸。一夜之间，这条蠕虫闪电般地自我复制，并向着整个 Internet 网络迅速蔓延，使美国 6000 余台基于 UNIX 的小型计算机和工作站受到感染和攻击，网络上几乎所有的机器都被迫停机，直接经济损失在 9000 万美元以上，莫里斯本人也因此受到了法律的制裁。从此以后，蠕虫也是一种病毒的概念被确立起来，而这种利用系统漏洞进行传播的方式也成为现在蠕虫病毒的主要传播方式。

蠕虫病毒的传染机理是利用系统漏洞通过网络进行复制和传播，传染途径是网络、电子邮件以及 U 盘、移动硬盘等移动存储设备。2006 年以来危害极大的熊猫烧香病毒就是蠕虫病毒的一种。蠕虫程序像生物蠕虫一样从一台计算机传染到另一台计算机，传播速度非

常快。

蠕虫病毒由两部分组成：一个主程序和一个引导程序。主程序一旦在机器上建立起来就会去收集与当前机器联网的其他机器的信息。它能通过读取公共配置文件并运行显示当前网上联机状态信息的系统实用程序而做到这一点，随后尝试利用前面所描述的那些缺陷在这些远程机器上建立其引导程序。

11.4.2 蠕虫病毒的传播及特点

1. 蠕虫病毒的特点

蠕虫病毒与一般的计算机病毒不同，它不采用将自身复制附加到其他程序中的方式来复制自己，因此在病毒中它也算是一个另类。蠕虫一般不采取利用 PE 格式插入文件的方法，而是通过复制自身在 Internet 环境下进行传播。病毒的传染能力主要是针对计算机内的文件系统，而蠕虫病毒的传染目标是 Internet 内的所有计算机，局域网条件下的共享文件夹、电子邮件 E-mail、网络中的恶意网页、大量存在着漏洞的服务器等都是蠕虫传播的良好途径。网络的发展也使得蠕虫病毒可以在几个小时内蔓延全球。而且蠕虫的主动攻击性和突然爆发性将使得人们手足无策。1988 年 Morris 蠕虫爆发后，Eugene H. Spafford 为了区分蠕虫和病毒，给出了蠕虫和病毒在技术角度的定义：计算机蠕虫可以独立运行，并能把自身的一个包含所有功能的版本传播到另外的计算机上；计算机病毒是一段代码，能把自身加到其他程序包括操作系统上，它不能独立运行，需要由它的宿主程序运行来激活它。

蠕虫病毒与一般的计算机病毒的区别如表 11-3 所示。

表 11-3 蠕虫病毒和一般的计算机病毒的区别

	一般的计算机病毒	蠕 虫 病 毒
存在形式	寄生	独立程序
复制机制	插入到宿主程序(文件)中	自身的复制
传染机制	宿主程序运行	系统存在漏洞
传染目标	主要针对本地文件	主要针对网络上其他计算机
触发传染	计算机使用者	程序自身
影响重点	文件系统	网络性能、系统性能
计算机使用者角色	病毒传播中的关键环节	无关
防治措施	从宿主程序中摘除	给系统打补丁

2. 蠕虫病毒的传播

蠕虫病毒的破坏性很强，部分蠕虫病毒不仅可以在 Internet 上兴风作浪，还把局域网当成它们的舞台。蠕虫病毒可以潜伏在基于客户机/服务器模式的局域网的服务器上的软件内，当客户机访问服务器并对有毒的软件实施下载后，病毒就神不知鬼不觉地从服务器上“复制”到客户机上了。

脚本病毒是很容易制造的，它们利用了 Windows 系统开放性的特点，特别是 COM 到 COM＋的组件编程思路，一个脚本程序能调用功能更大的组件来完成自己的功能。以 VB 脚本病毒(例如欢乐时光、I Love You、库尔尼科娃病毒、Homepage 病毒等)为例，它们都是把 vbs 脚本文件添在附件中，最后使用 *.htm.vbs 等欺骗性的文件名。

蠕虫病毒的实体结构包括未编译的源代码、已编译的链接模块、可运行代码、脚本、受感染系统上的可执行代码和信息数据。蠕虫病毒的基本程序结构为：

(1) 传播模块：负责蠕虫病毒的传播。

(2) 隐藏模块：侵入主机后，隐藏蠕虫病毒程序，防止被用户发现。

(3) 目的功能模块：实现对计算机的控制、监视或破坏等功能。

传播模块又可以分为 3 个基本模块：扫描模块、攻击模块和复制模块。

蠕虫病毒程序的一般传播过程为：

(1) 扫描：由蠕虫病毒的扫描功能模块负责探测存在漏洞的主机。当程序向某个主机发送探测漏洞的信息并收到成功的反馈信息后，就得到一个可传播的对象。

(2) 攻击：攻击模块按漏洞攻击步骤自动攻击步骤(1)中找到的对象，取得该主机的权限(一般为管理员权限)，获得一个 shell。

(3) 复制：复制模块通过源主机和新主机的交互将蠕虫病毒程序复制到新主机并启动。

蠕虫病毒工作方式流程图如图 11-4 所示。

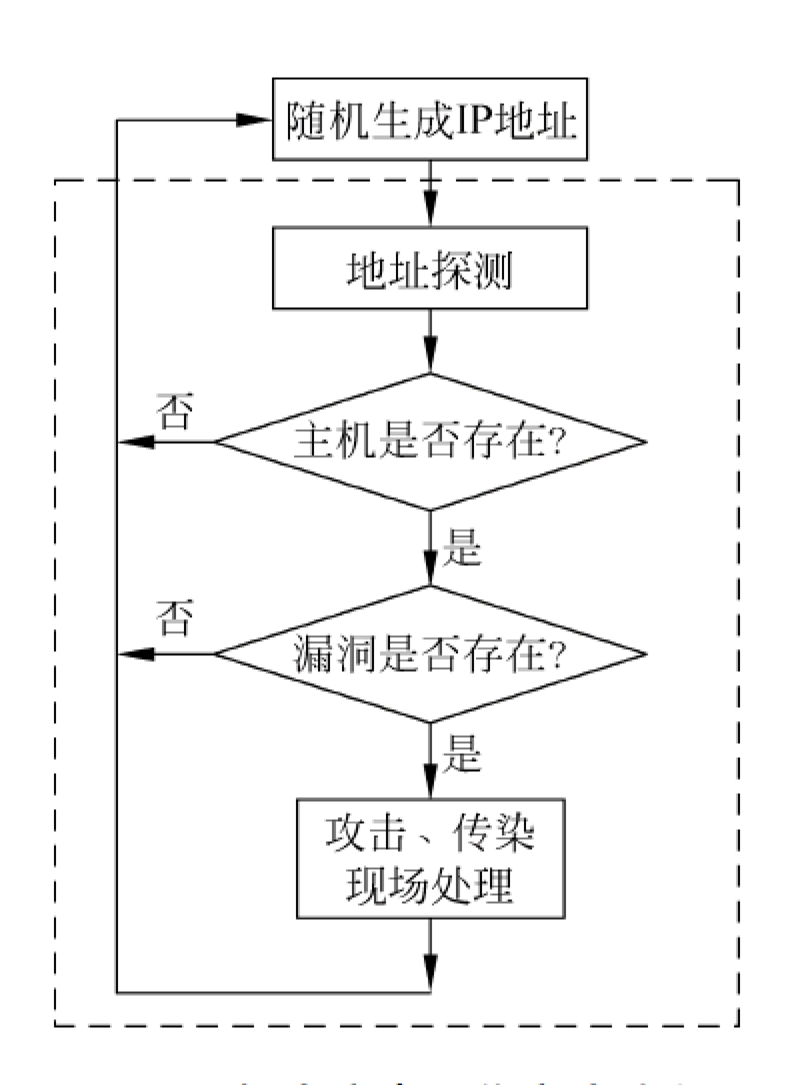


图 11-4　蠕虫病毒工作方式流程图

其中，有些蠕虫病毒随机生成 IP 地址，有些蠕虫病毒给出确定的地址范围，还有一些给出倾向性策略，用于产生某个范围内的 IP 地址。图 11-4 中虚线框内的所有工作可以在一个数据包内完成。可以看出，传播模块实际上是实现自动入侵的功能。因此蠕虫病毒的传播技术是蠕虫病毒技术的首要技术，没有蠕虫病毒的传播技术也就谈不上什么蠕虫病毒技术了。

11.4.3　常见蠕虫病毒介绍及防治方法

冲击波/振荡波病毒、SQL 蠕虫病毒、伪造源地址 DDoS 攻击、ARP 欺骗等是在宽带接入的网吧、企业、小区局域网内最常见的蠕虫病毒攻击形式。这几种病毒发作时会非常消耗局域网和接入设备的资源，造成用户上网很慢或者不能上网。

1. 冲击波/振荡波病毒

感染此类病毒的计算机和网络的故障现象有：

(1) 不断重新启动计算机或者莫名其妙地死机。

(2) 系统资源被大量消耗，导致 Windows 操作系统速度极慢。

(3) 中毒的主机大量发包阻塞网络，整个网络会迅速被这些攻击所形成的流量影响，形成 DoS 拒绝服务攻击。

(4) 造成局域网内所有用户的网速变慢直至无法上网。

局域网的主机在感染冲击波、振荡波及其变种病毒之后，会向外部网络发出大量的数据包，以查找其他开放了这些端口的主机进行传播，常见的端口有 TCP 135、TCP 139、TCP 445、TCP 1025、TCP 4444、TCP 5554、TCP 9996 以及 UDP 69 端口等。

快速查找此类病毒的方法是在 WebUI&#61664 上网监控页面查询当前全部上网记录，可以看到感染冲击波病毒的主机发出的大量 NAT 会话，特征有：

(1) 协议为 TCP,外网端口为 135/139/445/1025/4444/5554/9996 等。

(2) 会话中该主机有上传包,下载包往往很小或者为 0。

针对该类病毒的解决办法有:

(1) 将中病毒的主机从内网断开,杀毒,安装微软提供的相关的 Windows 补丁。

(2) 在安全网关上关闭该病毒向外发包的相关端口。

2. SQL 蠕虫病毒

SQL 蠕虫是一个能自我传播的网络蠕虫病毒,专门攻击有漏洞的微软 SQL Server TCP 1433 或 UDP 1434 端口,它会试图在 SQL Server 系统上安装自身并向外传播,进一步通过默认系统管理员 SQL Service 账户威胁远程系统。

SQL 蠕虫病毒是由一系列的 DLL、exe、bat 以及 js 文件构成,这些文件包含了一些 IP/端口扫描及密码盗取工具。它会将这些文件复制至受感染计算机上,并将 SQL 管理员密码改为由 4 个随机字母组成的字符串。

中毒的主机大量发包阻塞网络,整个网络会迅速被这些攻击所形成的流量影响,形成 DoS 拒绝服务攻击,造成局域网内所有用户网速变慢直至无法上网。

快速查找此类病毒的方法是在 WebUI&#61664 上网监控页面查询当前全部上网记录,可以看到感染 SQL 蠕虫病毒的主机发出的大量 NAT 会话,特征有:

(1) 协议为 TCP,外网端口为 1433;协议为 UDP,外网端口为 1434。

(2) 会话中有上传包,下载包往往很小或者为 0。

针对该类病毒的解决办法有:

(1) 将中病毒的主机从内网断开,杀毒,安装微软提供的相关的 SQL Server 补丁。

(2) 在安全网关上关闭该病毒的相关端口。

3. 伪造源地址 DDoS 攻击

伪造源地址攻击中,黑客机器向受害主机发送大量伪造源地址的 TCP SYN 报文,占用安全网关的 NAT 会话资源,最终将安全网关的 NAT 会话表占满,导致局域网内所有用户无法上网。

快速查找此类病毒的方法是在 WebUI&#61664 上网监控页面中查询,看到"IP 地址"一栏里面有很多不属于该内网 IP 网段的用户。如果安全网关接收到某用户发送的海量数据包,但安全网关发向该用户的数据包很小,依此可以判断该用户可能在进行伪造源地址攻击。

针对该类病毒的解决办法有:

(1) 将中病毒的主机从内网断开,进行杀毒。

(2) 在安全网关配置策略时只允许内网的网段连接安全网关,让安全网关主动拒绝伪造的源地址发出的 TCP 连接。

4. ARP 欺骗攻击

当局域网内某台主机运行 ARP 欺骗的木马程序时,会欺骗局域网内所有主机和安全网关,让所有上网的流量必须经过病毒主机。其他用户原来直接通过安全网关上网现在转由通过病毒主机上网,切换的时候用户会断一次线。

切换到病毒主机上网后,如果用户已经登录了某游戏服务器,病毒主机又经常伪造断线

的假象,在用户多次重新登录服务器时,病毒主机就会乘机进行盗号。

由于 ARP 欺骗的木马程序发作时会发出大量的数据包导致局域网通信拥塞以及自身处理能力的限制,用户上网速度越来越慢。当 ARP 欺骗的木马程序停止运行时,用户会恢复从安全网关上网,切换过程中用户会再断一次线。

快速查找此类病毒的方法是在 WebUI系统状态 系统信息系统历史记录中,看到大量如下的信息:

```
MAC SPOOF 192.168.16.200
MAC Old 00:01:6c:36:d1:7f
MAC New 00:05:5d:60:c7:18
```

这个消息代表了用户的 MAC 地址发生了变化。在 ARP 欺骗木马开始运行的时候,局域网所有主机的 MAC 地址更新为病毒主机的 MAC 地址(即所有信息的 MAC New 地址都一致为病毒主机的 MAC 地址)。

同时,在安全网关 ARP 表中看到所有用户的 MAC 地址信息都一样,或者在用户统计中发现所有用户的 MAC 地址信息都一样。

通常采用双向绑定的方法解决并且防止 ARP 欺骗:

(1) 在 PC 上绑定安全网关的 IP 和 MAC 地址。

(2) 在安全网关上绑定用户主机的 IP 和 MAC 地址。

11.4.4 防范蠕虫病毒的安全建议

(1) 蠕虫病毒的一般防治方法是使用具有实时监控功能的杀毒软件。

(2) 防范邮件蠕虫的首要方法是提高安全意识,勤打补丁,定时升级杀毒软件和防火墙。对于网络管理员来说,还要对系统定期备份,尤其是多机备份,防止意外情况下的数据丢失;对于局域网用户,可以在 Internet 入口处安装防火墙,对邮件服务器进行监控;对于个人用户,上网要尽量选择一些大型的门户网站,对于来历不明的电子邮件,尤其是附件,最好不要打开。另外,机器的安全设置可以设置得高一些,例如 IE 的安全级别可以设置为中,将其中所有 ActiveX 插件以及 Java 相关控件全部选择禁用。不过,这样会造成在网页浏览过程中个别含有 ActiveX 的网站无法浏览,影响用户体验。此外,还可以在 Office 里面禁用宏等。

(3) 对于某些网络蠕虫病毒通过调用系统中已经编译好的带有破坏性的程序来实现这一功能,用户可以把本地带有破坏性的程序的名字更改,例如将 format.com 改成 fmt.com,从而使得病毒的编辑者无法调用本地命令来实现这一功能。

(4) 由于蠕虫病毒大多是用 VBScript 脚本语言编写的,而 VBScript 代码是通过 Windows Script Host 来解释执行的,如果将 Windows Script Host 删除,就再也不用担心这些用 VBS 和 JS 编写的病毒了。从另一个角度来说,Windows Script Host 本来是被系统管理员用来配置桌面环境和系统服务,实现最小化管理的一个手段。但对于大部分的一般用户而言,WSH 并没有多大用处,因此可以禁止 Windows Script Host。可以在 C:\Windows\System32 目录下,找到 WScript.exe 等脚本程序的系统支持文件,更改其名称或者干脆删除该文件。后一种做法有一定的副作用,如果删除脚本程序的系统支持文件的话,网页的 JS 和 WS、VBS 等脚本将不能再执行,因此要慎用。

11.5 恶意软件的危害

恶意软件的特征使得它很容易流行,也必定给用户带来诸多的不便,造成用户计算机速率和存储空间上的影响,同时会给整个网络以及信息安全带来危害,影响用户在网络上的正常活动。恶意软件的危害大致有3个方面。

1. 从整体上降低计算机运行效率

恶意软件中的广告软件、网络插件等,都具有强制安装、难以卸载、自动启动等特征。这些软件在强制安装后,通常会在用户使用计算机时自动启动,占用系统的一些进程和一定的内存空间,影响用户的正常操作,甚至导致计算机瘫痪。

2. 妨碍人们的正常网络活动

恶意软件中通常有些诸如广告软件和浏览器劫持的恶意程序,这些软件并不能直接危害到用户的信息安全,但其强制性却给用户带来了许多不便。广告软件通常会在浏览器中添加自己的插件,频繁弹出广告窗口,占用较多系统资源,从而影响用户浏览网页等网络活动的速度。浏览器劫持则会在浏览网页的时候强制浏览指定的网页,大大影响用户的操作。这一类软件给操作和使用带来极大不便,影响使用网络的效率。

3. 危害网络信息的安全

恶意软件中的间谍软件、行为记录软件等会获取计算机用户的资料等信息,甚至可以远程控制用户计算机,严重侵犯了用户的信息安全。2007年微软的报告指出木马下载和投放程序以及相关的恶意代码数量增长了500%。近年来随着网络的普及,类似的恶意软件越来越多。用户的各种账户尤其是银行系统的账户受到严重威胁,造成个人信息的泄露以及财产损失。

11.6 恶意软件防范与清除

11.6.1 恶意软件防范

计算机用户必须在使用计算机的过程中加强安全防范意识,并利用掌握的计算机知识尽可能多地排除系统安全隐患,力求将其阻挡在系统之外。

1. 加强系统安全设置

1) 及时更新系统补丁

在操作系统安装完毕后,尽快访问微软站点下载最新的Service Pack和漏洞补丁,并将计算机的"系统更新"设置为自动,最大限度地减少系统存在的漏洞。

2) 严格账户管理

使用Guest账户,把Administrator账户改名,删除所有的测试账户、共享账户等。不同的用户组设置不同的权限,严格限制具有管理员权限的用户数量,确保不存在密码为空的账户。

3）关闭不必要的服务和端口

禁用一些不需要的或者存在安全隐患的服务。如果不是应用所需就关闭远程协助、远程桌面、远程注册表、Telnet 等服务。对于个人用户来说，系统安装中默认的有些端口是没有必要开放的。在\system32\drivers\etc\services 文件中有知名端口和服务的对照表可供参考，用户可以根据应用需要选择开放端口，关闭其他不用的端口。

2. 养成良好的计算机使用习惯

在使用计算机过程中，不要随意打开不明网站，很多恶意软件都是通过恶意网站进行传播的。到知名正规网站下载软件，安装软件时要细看慢点，在安装时注意选择要安装的插件，尽量减少异常。禁用或限制使用 Java 程序及 ActiveX 控件，这些程序的脚本中往往会含有恶意代码，给用户带来不便。

计算机的发展是迅速并不断变化的，这就要求用户的计算机知识也必须顺应时代的变化而变化。补充知识能够使用户对计算机的认识逐渐深入，最大限度地降低恶意软件所带来的影响。

3. 增强法律保护意识

恶意软件会给用户带来不便，甚至侵犯用户的权益。一些恶意软件被用来进行不正当竞争，侵犯他人的合法权益。这时就需要用户拿起法律武器保护自己的合法权益，用法律维护公平，减少恶意软件的危害。

11.6.2　恶意软件清除

感染恶意软件后，计算机通常会出现运行速度变慢、浏览器异常、系统混乱甚至系统崩溃等问题。因此掌握恶意软件的清除方法，对于广大计算机用户来说是十分必要的。

1. 手工清除

如果发觉自己的计算机感染了少量恶意软件，可以尝试用手工方法将其清除。具体方法如下：

(1) 重启计算机，开机按 F8 键，选择进入安全模式。

(2) 删除浏览器的 Internet 临时文件、Cookies、历史记录和系统临时文件。

(3) 在“控制面板”→“添加或删除程序”中查找恶意软件，如果存在将其卸载。找到恶意软件的安装目录，将其连同其中的文件一并删除。

(4) 在“运行”中输入“regedit”，进入注册表编辑器，在注册表中查找是否存在含有恶意软件的项、值或数据，如果存在，将其删除。

以上 4 步完成后，重启计算机进入正常模式，通常恶意软件即可被清除。

2. 借助专业清除软件

随着恶意软件技术越来越高，使用手工的方法已很难彻底清除它们，可以借助一些专业的软件来进行清除。

目前，Internet 上可供查杀恶意软件的工具多达数十款，需要注意的是最好在安全模式下运行这些清除软件，查杀才能更为彻底。

11.7 恶意软件的发展趋势

随着各类恶意软件相关技术的日益成熟,恶意软件的发展出现了一些值得关注的新趋势,了解和把握这些趋势将有利于针对性地制定相应的防护措施。

1. 由任意传播转向定量传播

在2004年之前的数年间,Worm和Bot在传播时采用的主要是任意传播策略。对于其搜索到的任何可利用、入侵的目标系统都进行感染,将自身传播到目标系统中。根据多家安全服务公司及防病毒软、硬件厂商的监控数据,从2004年5月至今,已极少监测到类似前几年CodeRed Worm、SlammerWorm等大规模传播的Worm/Bot,而因恶意软件导致的各种安全问题发生频率却越来越高。综合防病毒软、硬件厂商对相关数据的分析结果表明,当前的Worm/Bot传播策略发生了明显的调整,已由任意传播策略转向定量传播策略,即Worm和Bot在传播过程中仅感染满足特定条件的目标系统,并且当成功感染目标系统的数量达到一定值后,则不再继续传播。根据Kaspersky Labs的统计分析,这一临界数量在5000～10000之间。

任意传播策略的特征是传播速度快、被感染的目标系统数量大。这两个特征导致了以任意传播策略传播的Worm/Bot很容易被防病毒软、硬件厂商采集到其样本,进而提取其Signature(特征码)并更新病毒库,从而使得普通用户能及时、有效地清除这些Worm/Bot并阻止其在网络上的继续传播。

定量传播策略的特征是传播速度快,但被感染的目标系统数量相对而言非常少。因此,基于此策略传播的Worm/Bot很难被防病毒软、硬件厂商采集到样本,安装了防病毒软、硬件系统的普通用户即使及时更新病毒库,也无法查出系统中已经存在的Worm/Bot,无法过滤进入系统中的Worm/Bot。

2. 存活能力显著提高

(1) 在传播过程中,通过Polymophism(变形)技术处理后的恶意软件不再具有固定的Signature,从而可有效对抗当前防病毒软、硬件的过滤和查杀。精心编写的恶意软件甚至可以将Polymophism引擎包含在其自身内部,使得每次传播都具有不同的Signature。

(2) 利用Rootkit实现对进程、线程、通信连接、通信数据、通信目的地的隐藏,同时结合Firewall/ IDS Bypass(防火墙旁路)、Firewall/ IDS Disable(禁用防火墙/入侵检测系统)、降低系统安全等级设置等技术,可有效防止在被入侵的目标系统内运行着的恶意软件被检测到。

综合运用以上措施后,当前恶意软件能有效躲避当前广泛应用的、传统的、基于Border Firewall+DMZ IDS+Personal Firewall+Antivirus Software结构的Defense in Depth(深度防御)体系的检测,存活能力显著提高。

3. 应对基于系统缺陷的攻击响应时间迅速缩短

Zero-Day Exploit是指缺陷被发现后立即出现的漏洞利用程序。Zero-Day Exploit的不断涌现标志着发现系统缺陷到漏洞利用程序出现的间隔时间大大缩短,从而导致应对基

于系统缺陷的攻击的响应时间迅速缩短。当出现 Zero-Day Exploit 时，软件厂商往往尚未发布相应补丁。即使有补丁发布，多数用户也无法保证在补丁出现时立即给系统打上补丁，从而导致大量有缺陷的系统暴露于 Zero-Day Exploit 的威胁之下。

4. 传播途径多样化

众多 P2P(Peer-to-Peer，点到点）模式以及 IM(Instant Messaging，即时通信)类网络应用程序的迅猛发展，为恶意软件提供了更便捷、更快速的传播途径。

5. 更易用、功能更强大

当前恶意软件的发展使得入侵者只需很少的专业知识即可掌握其用法。同时，综合运用各种恶意软件相关技术的混合型恶意软件的功能越来越强大，一旦入侵成功，即可实现对目标系统的完全控制。

6. 恶意软件侵袭 3G 网络

近年来，3G 技术和智能手机的发展势头异常迅猛。虽然手机病毒在目前还没有非常适合的传染源、传染途径和传播目标，短时间内并不会大规模地爆发。但是，随着 3G 手机的推广和智能手机的普及，手机病毒可能带来的问题不容忽视。从目前的形势来看，在移动通信产业中的金融诈骗会越来越多，并且同样面临着与 PC 终端相同的恶意软件风险。

安全厂商 McAfee 在全球移动大会(Mobile World Congress)上发表报告称，半数手机制造商报告了恶意软件感染、语音或文字垃圾消息攻击和其他安全事件。同时还指出，70%的手机制造商认为手机安全对于未来的发展至关重要，手机安全保护是一项重要举措。

某安全厂商曾检测到一种攻击 Symbian 系统的新恶意程序。该程序的攻击目标是一家印度尼西亚移动电话运营商的用户。该木马用 Python 编写，会发出短信，通过一连串简短号码的指示，将用户账户里的部分资金转移到另一个属于网络犯罪分子的账户中。目前，已知的 5 个 SMS. Python. Flocker 变种木马，转移金额的范围从 0.45 美元到 0.90 美元不等。如果网络犯罪分子设法让木马感染了大量手机的话，那么转移到犯罪分子手机账户的金额数目将是相当可观的。

思　考　题

(1) 恶意软件的定义是什么？列举尽可能多的恶意软件种类。

(2) 当前恶意软件表现出什么样的发展趋势？

(3) 根据恶意软件命名规则，解读 Win32. Happy99. Worm 的含义。

(4) 木马是一种特殊的恶意软件，它的本质特征是什么？

(5) 简述恶意软件的基本特征。

(6) 了解恶意软件的危害及常用防范手段。

参考文献

[1] 王健.“流氓软件”互联网的黑社会.法律与生活,2006,(10).

[2] 周强. 恶意软件法律法规之探析.中国政法大学民商法学,2008.

[3] 那罡. 恶意软件侵袭3G. 中国计算机报,2009. 3. 16,(034).

[4] Baecheer,P. Koetter,M. Holz,T. Dornseif,Freiling. The Nepenthes Platform: An Efficient Approach To Collect Malware. In Recent Advances in Intrusion Detection (RAID),2006.

[5] Kirda,E. Kruegel,C. Banks,G. Vigna,and Kemmerer. Behavior-based spyware Detection. In 15th Usenix Security Symposium,2006.

[6] 卢浩,胡华平,刘波. 恶意软件分类方法研究. 计算机应用研究,2006.

[7] 余前佳. 恶意软件的特征、危害性及其防范与清除方法土国.资源信息化,2006,(6):39～41.

[8] 刘青风,李敏. 恶意软件的防护方法探讨.软件导刊,2007,(7):96～98.

[9] 系统安全:Windows 系统安全设置方法,赛迪网,2006,(8).

[10] 尚志红,刘羽燕,熊江河. 恶意软件隐私侵权问题研究. 经济研究导刊,2011,(7).

[11] 周再红,郭广军,申东亮. 间谍软件攻击分析与对策研究. 湖南人文科技学院学报,2004,(6).

[12] 周琳洁. 网络环境下恶意软件的特征、危害和防范研究. 内蒙古科技与经济,2011,(6).

[13] Malware. http://en.wikipedia.org/wiki/Malware.

[14] 王建锋,钟玮,杨威. 计算机病毒分析与防范大全. 北京:电子工业出版社,2011.

第12章　网络入侵与取证

本章学习目标

随着计算机的普及，网络应用越来越广泛，网络安全问题也日益严重，其中的网络入侵问题也越来越受到专业人士的重点关注。计算机网络的入侵检测，指的是对计算机网络及其整体系统的检测，检查是否存在违反安全原则的事故发生。网络的入侵取证系统是对网络防火墙合理的补充，是对系统管理员安全管理能力的进一步扩展，有助于提高网络安全基础结构的完整性。

通过对本章的学习，应掌握以下内容：

(1) 导致网络脆弱的因素。

(2) 攻击者实施网络入侵的常用方法及防范方法。

(3) 入侵检测系统的原理、结构和流程。

(4) 计算机取证的一般步骤和取证模型。

12.1　网络的概念

12.1.1　网络

通常情况下，将网络中的一个单一计算系统称为一个节点，计算机称为主机，两台主机之间的连接称为链路。网络计算由许多的用户、通信介质、可见主机和通常对终端用户不可见的系统组成。通过直接操作终端、工作站或计算机完成用户与网络系统的通信。工作站是终端用户的计算设备，一般都配置有功能强大的处理器、大量的内存和存储空间，以实现一些复杂的数据处理工作。系统是一个处理器的集合，通常具有比工作站更强的处理能力及更大的存储空间。

12.1.2　网络的特征

网络具有以下几个典型的特征：

(1) 匿名性。网络隐藏了通信者绝大多数的特征，例如相貌、声音等。

(2) 自动性。对于某些网络中的一个具体通信而言，通信的一端或两端都可能是由机器自动完成的，需要较少的人工管理。

(3) 远程性。网络连接的端点之间距离通常很远。但是由于如今的通信速度已经足够快速，通信者通常感觉不出另一端的站点与己端的距离。

(4) 透明性。用户不仅不能判断远程主机的距离，也不能区分连接到的节点本身的计算系统与计算能力。

(5) 路由的相异性。为了维持与提高网络的可靠性与性能，两个端点间的通信路由通

常都是动态分配的。几秒钟前通过网络传输的路径可能就与下一次传输的路径有很大的不同。

12.1.3　网络的类型

1. 局域网(LAN)

局域网的覆盖范围较小,一般局限于一栋建筑内。通常,一个局域网连接着若干台小型计算机、打印机和一些专用文件存储设备。局域网的主要优点在于它的所有用户可以很方便地共享数据、程序及设备。

局域网的特征有:

(1) 规模小。通常情况下,共享同一个局域网的用户人数较少,其覆盖范围也不超过 3km。

(2) 局部控制。都由一个组织统一管理。

(3) 物理保护。局域网常被规定在一个公司或者其他组织内部使用。因此,局域网以外的恶意访问者通常很难访问到局域网内的设备。

(4) 有限范围。很多局域网支持单一的组、部门、楼层。所完成的功能仅覆盖一个很窄的范围。

2. 广域网(WAN)

广域网和局域网在规模、距离和控制或拥有关系等方面都有较大的区别,一些广域网处于密切的控制和维护之下,以实现高度的逻辑和物理隔离,而其他广域网仅仅是为了使用方便而连接在一起。因此在同一个广域网上的主机可能属于彼此相隔几千里、相互独立的多个公司,这样做的目的是为了共享硬件设备等资源以降低生产成本。

广域网的典型特征有:

(1) 单一控制。一个广域网通常由一个组织机构负责管理,由它决定谁可以加入该网络并使用其中的资源。

(2) 覆盖范围大。一个广域网的覆盖范围通常比一个局域网的大。

(3) 物理上暴露。大多数广域网使用公共通信介质,而这些公共介质相对来说要暴露一些。但如果很多用户共享这些介质,则有助于保护这些用户相互之间的私有通信。

其他网络类型还包括校园网(CAN)和城域网(MAN)。一个校园网通常是由一个大学或一家公司等组织单独进行控制的。一个城域网通常覆盖一座城市,由该区域内的一个机构提供通信服务。校园网、城域网和广域网覆盖范围都可能较为广泛,相互之间的区别不是很严格,但是都介于局域网和 Internet 之间。

3. 互联网(网际网)

互联网是由众多网络构成的网络。最典型的互联网是 Internet,它由 Internet 协会进行松散的控制,该协会制定了一些公平活动的基本规则,以保证所有用户都能接受公平的服务,并且它支持标准协议,以实现用户之间的通信。

Internet 的特征有:

(1) 联合。用户可能通过商业组织或政府组织网络进入 Internet,也可能通过按月付费来连接到 Internet。但即使是网络服务提供商,都难以对 Internet 给出一个适当的描述。

(2) 巨大。Internet 连接了几千个网络,每天都有新的主机加入,没有人能计算出 Internet 到底有多大。

(3) 异构。可能每一种商业硬件和软件中都至少有一个产品连接到了 Internet 上。大多数多用户操作系统都可以支持对 Internet 的访问。

(4) 物理上和逻辑上暴露。由于没有统一的访问控制,实际上任何攻击者都可以访问 Internet,并且由于连接设备的复杂性,攻击者可以获取网上的任何资源。

12.2　网络面临的威胁

12.2.1　导致网络脆弱的因素

1. 匿名性

攻击者无须与被攻击的系统、该系统的管理员或者任何用户进行直接接触就可以实施攻击。攻击者可能利用很多其他主机来实施攻击,从而隐藏攻击的源头。

2. 攻击点多

攻击者可以同时攻击多个目标,也可以在多个地点发动攻击。独立的计算系统是一个自包含的单位,本机的访问控制机制保证了在该处理机上数据的机密性。然而,当用户要取得远程网络主机上存储的文件时,该文件数据要经过很多主机的传递才能到达该用户。因此,用户必须依赖于所有这些系统的访问控制机制。在一个大型网络中,容易受到攻击的点是相当多的。

3. 共享性

由于网络允许节点间共享资源和分担负载,所以潜在的访问联网系统的用户比单机用户还多。还有更多的系统可以访问网络资源。

4. 系统的复杂性

网络是由两个或者更多可能不相同的操作系统组成的,攻击者可以利用计算机强大的计算能力,将其部分攻击计算工作交给攻击目标完成,从而能够大大提高攻击能力。

5. 未知边界

网络的可扩展性意味着网络的边界是不确定的。一台主机可能接入了多个不同网络。提供广泛的访问能力是网络的一个优势,然而,怀有恶意的未知的或者不受控制的用户确是网络安全的不利因素。一旦允许新主机加入网络,就有可能出现类似的问题。

6. 未知路径

从一台主机到另一台主机可能存在着多条路径。网络用户很难控制他们所发消息的传递路由路径。消息可能经过提供了相应安全措施的主机,也有可能经过没有提供安全措施、存在隐患的主机。

以上这些网络特征明显地增加了网络所面临的安全危险,表 12-1 列出了一些常见的网络攻击方法。

表 12-1 常见的网络攻击方法

目标	攻击方法
破坏机密性	协议缺陷、偷听、被动窃听、误传、网络内暴露、流量分析、窃取 Cookies
破坏完整性	协议缺陷、主动窃听、假冒、伪造消息、噪声、DNS 攻击
破坏可用性	协议缺陷、连接洪泛、DNS 攻击、流量重定向、分布式拒绝服务攻击
获取信息	端口扫描、社会工程学、侦查、OS 与应用软件的特征
破坏鉴别	假冒、猜测、窃听、欺骗、会话劫持、中间人攻击
利用编程缺陷	缓冲区溢出、寻址错误、参数修改、恶意活动代码、恶意代码、恶意输入代码

12.2.2 搜集网络漏洞信息的常用方法

实施网络入侵的攻击者可采取各种不同的攻击手段，但是在采取攻击行动之前，都会做好充分的调查和计划，攻击者往往先通过网络查找相关目标尽可能多的资料。

1. 端口扫描

端口扫描技术是一项探测本地或远程系统端口开放情况的策略和方法，通过向目标主机的 TCP/IP 服务端口发送探测数据包，并记录目标主机的响应。通过分析响应来判断服务端口的状态是打开还是关闭，就可以得知端口提供的服务或信息。对接收到的数据进行分析，进一步发现目标主机的某些内在弱点。

TCP/UDP/IP 常被攻击者利用的端口号如表 12-2 所示。

表 12-2 易被攻击者利用的端口号

端口	服务	说明
21	FTP	FTP 服务器所开放的端口，用于上传、下载。最常见的用于寻找打开 anonymous 的 FTP 服务器。木马 Doly Trojan、Fore、Invisible FTP、WebEx、WinCrash 和 Blade Runner 所开放的端口
23	Telnet	远程登录，入侵者扫描这一端口获取机器运行的操作系统信息。木马 Tiny Telnet Server 开放的端口
53	Domain Name Server (DNS)	DNS 服务器所开放的端口，入侵者可能是试图进行区域传递(TCP)，欺骗 DNS(UDP)或隐藏其他通信。因此防火墙常常过滤或记录此端口
69	Trival File Transfer	许多服务器与 BOOTP 一起提供这项服务，便于从系统下载启动代码。但是它们常常由于错误配置而使入侵者能从系统中窃取任何文件
79	Finger Server	入侵者用于获得用户信息，查询操作系统，探测已知的缓冲区溢出错误，回应从自己机器到其他机器 Finger 扫描
80	HTTP	用于网页浏览。木马 Executor 开放此端口
99	Metagram Relay	后门程序 ncx99 开放此端口
109	Post Office Protocol-Version3	POP3 服务器开放此端口，用于接收邮件，客户端访问服务器端的邮件服务。POP3 服务有许多公认的弱点。关于用户名和密码交换缓冲区溢出的弱点至少有 20 个，入侵者可以在真正登录前进入系统。成功登录后还有其他缓冲区溢出错误
177	X Display Manager Control Protocol	许多入侵者通过它访问 X-Window 操作台，它同时需要打开 6000 端口

续表

端口	服　务	说　明
553	CORBA-IIOP（UDP）	使用 Cable Modem、DSL 或 VLAN 将会看到这个端口的广播。CORBA 是一种面向对象的 RPC 系统。入侵者可以利用这些信息进入系统
1080	SOCKS	这一协议以通道方式穿过防火墙，允许防火墙后面的人通过一个 IP 地址访问 Internet。理论上它应该只允许内部的通信向外到达 Internet。但是由于错误的配置，它会允许位于防火墙外部的攻击穿过防火墙。WinGate 常会发生这种错误，在加入 IRC 聊天室时常会看到这种情况

攻击者通过端口扫描能够了解目标系统 3 个方面的内容：

(1) 目标系统上有哪些标准端口或者服务正在运行并响应请求。

(2) 目标系统上安装了哪种操作系统。

(3) 当前都有哪些应用服务在提供服务及其版本。

这些信息都可能在目标机器毫无察觉的情况下被攻击者获得，端口扫描过程几乎不会引起用户的注意。

端口扫描技术根据端口连接的方式主要可分为全连接扫描、半连接扫描、秘密扫描和其他扫描。

1) 全连接扫描

全连接扫描是 TCP 端口扫描的基础，现有的全连接扫描有 TCP connect 扫描和 TCP 反向 ident 扫描等。其中 TCP connect 扫描的实现原理为：扫描主机通过 TCP/IP 协议的三次握手与目标主机的指定端口建立一次完整的连接；连接由系统调用 connect 开始；如果端口开放，则连接将建立成功；否则若返回 −1，则表示端口关闭。

2) 半连接扫描

若端口扫描没有完成一个完整的 TCP 连接，在扫描主机和目标主机的指定端口建立连接时候只完成了前两次握手，在第三步时，扫描主机中断了本次连接，使连接没有完全建立起来。这样的端口扫描称为半连接扫描，也称为间接扫描。现有的半连接扫描有 TCP SYN 扫描和 IP ID 头 dumb 扫描等。

3) 秘密扫描

端口扫描有可能被在端口处所监听的服务器日志记录。这些服务器看到一个没有任何数据的连接进端口，就记录一个日志错误。而秘密扫描是一种不被审计工具所检测的扫描技术。现有的秘密扫描有 TCP FIN 扫描、TCP ACK 扫描、NULL 扫描、XMAS 扫描、TCP 分段扫描和 SYN/ACK 扫描等。

4) 其他扫描

(1) FTP 反弹攻击

FTP 反弹攻击是扫描主机通过使用 port 命令，探测到 USER-DTP(用户端数据传输进程)正在目标主机上的某个端口侦听的一种扫描技术。

(2) UDP ICMP 端口不可到达扫描

扫描主机发送 UDP 数据包给目标主机的 UDP 端口，等待目标端口不可到达的 ICMP 信息。若这个 ICMP 信息及时接收到，则表明目标端口处于关闭状态；若超时也未能接收

到端口不可到达的 ICMP 信息，则表明目标端口可能处于监听状态。

端口扫描工具种类非常多。nmap scanner 就是一款有用的工具。给定一个地址之后，nmap 就会报告其所有开放的端口、支持的服务以及提供服务的后台邮件收发程序的所有者。著名的商业扫描器有 Nessus、CyberCop Scanner、Secure Scanne、Internet Scanner 等。

2. 服务识别

扫描端口的目的是为了获取目标主机提供的服务信息，通过检测服务器开放的端口号，可以参照 RFC1700 标准推断出目标主机提供的服务。

主要有两种服务识别方法：

(1) 对于主动提供旗标信息或握手信息的服务可以使用 Netcat 尝试与目标的该端口直接建立连接，根据返回的信息作出初步判断。能提供该类服务的服务器通常称为主动式 Server。

(2) 还有一类服务需要客户端首先发送一个命令，然后再作出相应。要判断这样的服务，必须首先猜测服务类型，然后模仿客户端发送命令，等待服务器的回应。

根据服务器的回应参照 RFC1700 标准可以推断出目标主机提供的服务类型。

一般步骤如图 12-1 所示。

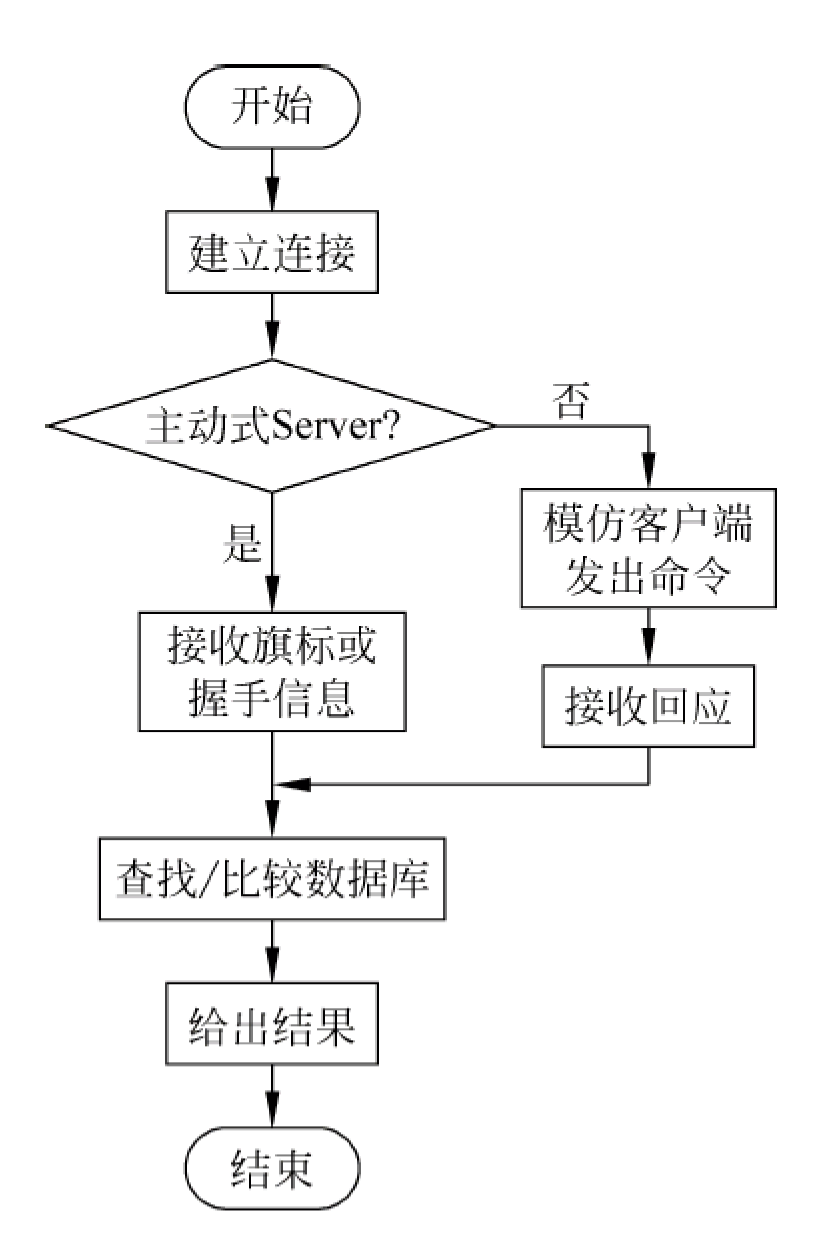


图 12-1 服务识别流程图

上述方法在以下 3 种情况下可能出现错误：

(1) 该目标主机将某服务故意开设到了非标准端口。

(2) 该目标主机开设了 RFC1700 中未定义的服务。

(3) 该目标主机被安置了后门程序。

3. 获取操作系统与应用程序特征

前文所述的端口扫描可以为攻击者提供目标的很多特殊信息，例如，攻击者可以用它发现某个系统的 80 端口是开放的，并支持用于网页浏览的 HTTP 协议。但是，攻击者可能还需要获取一些更为重要的信息，从而决定利用目标主机的哪一个漏洞实施攻击。

虽然网络协议都有统一的标准，然而开发商却是独立的，每一个开发商的应用实现的代码都各不相同。因此，在他们对标准进行解释和实现的时候总会有一些细微的变化。开发软件的不同版本都可能有不同的序号、TCP 标志以及新的可选项。一个系统如何进行响应也能泄漏系统的类型及其版本。将这些特征性信息称为操作系统或者应用程序的特征，这些特征为攻击者实施攻击行为提供了准备信息。

例如，nmap scanner 不仅能完成端口扫描，还会分析并猜测目标操作系统的相关情况，并向攻击者报告其猜测结果。总而言之，尽可能多地并且尽可能准确地获取目标系统的相关信息，有助于提高攻击者攻击行为成功的概率。操作系统探测方法比较如表 12-3 所示。

表 12-3　操作系统探测方法比较

方　法		优　点	缺　点
利用系统和服务的 Banner		简单、快速、有效	不太可靠，有的情况下无法获得 Banner 信息，有时可能被 Banner 信息欺骗
主动特征探测	ICMP 响应分析	准确性较高	防火墙阻塞 UDP 或 ICMP 等协议时不太可靠
	TCP 报文响应分析	需要一个打开的 TCP 端口、一个关闭的 TCP 端口和一个关闭的 UDP 端口，准确性高	在有防火墙阻挡的情况下，可能只有一个开放的 TCP 端口，这时准确性大大降低
	TCP 报文延时分析	只需要一个打开的 TCP 端口	速度慢
被动特征探测		不易被发现，主要被入侵者使用	分析数据更加复杂

12.2.3　网络入侵的常用方法及防范措施

入侵是指企图对计算机系统造成危害的行为。入侵企图或威胁可以被定义为未经授权蓄意尝试访问信息、篡改信息、使系统不可靠或不能使用，或者是指有关试图破坏资源的完整性、机密性及可用性的活动。入侵分为 6 种类型：

(1) 尝试性闯入(Attempted break-in)。

(2) 伪装攻击(Masquerade attack)。

(3) 安全控制系统渗透(Penetration of the security control system)。

(4) 泄露(Leakage)。

(5) 拒绝服务(Denial of service)。

(6) 恶意使用(Malicious use)。

攻击者入侵网络主要采取的方式有：拒绝服务攻击、协议欺骗攻击、密码猜测攻击、木马程序攻击、Web 欺骗攻击、邮件炸弹、缓冲区溢出攻击、Windows 系统漏洞攻击、SQL 注入、UNIX 系统攻击等。

1. 拒绝服务攻击(Denial of Service，DoS)

拒绝服务攻击是目前比较有效而又比较难以防御的一种网络攻击手段，其攻击目的是占用过多的服务资源，使服务器不能为正常访问的用户提供服务。因此，DoS 对一些紧密依靠 Internet 开展业务的企业和组织带来了致命的威胁。

拒绝服务攻击技术有多种，例如 SYN Flood、UDP Flood、Ping of Death、TearDrop、Land Attack、IP Spoofing DoS。其中 SYN Flood 是最常见又最容易被利用的一种 DoS 攻击形式，它利用 TCP 三次握手协议的缺陷，向目标主机发送大量的伪造源地址的 TCP SYN 报文，目标主机接收到报文后分配必要的资源，然后向源地址返回 SYN＋ACK 包，并等待源端返回 ACK 包。由于源地址是伪造的，所以源端永远都不会返回 ACK 报文，受害主机继续发送 SYN＋ACK 包，并将半连接放入端口的积压队列中，虽然一般的主机都有超时机制和默认的重传次数，但是由于端口的半连接队列的长度是有限的，如果不断地向受害主机发送大量的 TCP SYN 报文，半连接队列很快就会填满，服务器拒绝新的连接，将导致该端口无法响应其他机器进行的连接请求，最终耗尽目标主机的资源，如图 12-2 和图 12-3 所示。

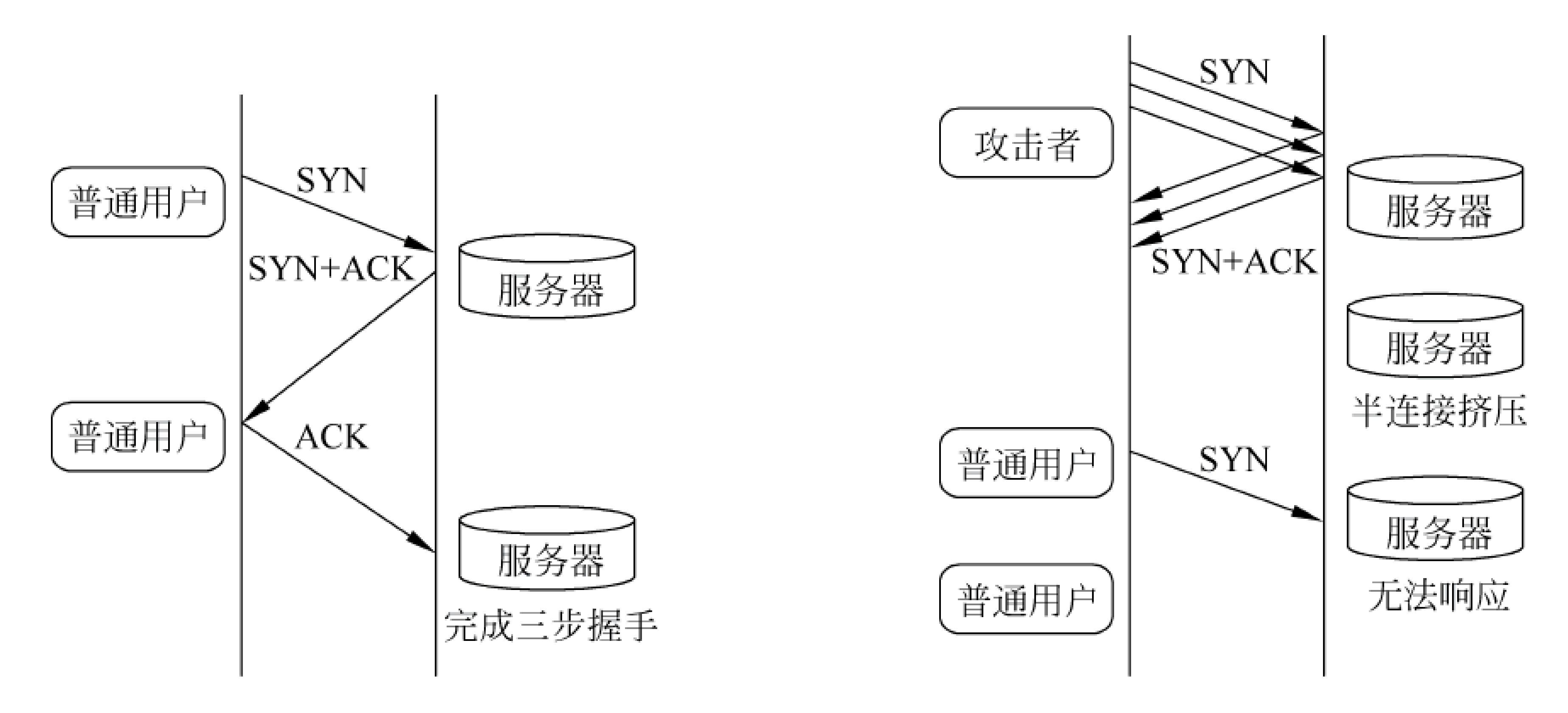


图 12-2 用户与服务器正常连接 **图 12-3 SYN FLOOD**

配置路由器、防火墙、入侵检测系统或入侵防御系统可以抵御常见的 DoS 攻击。要彻底杜绝 DoS 攻击,最好的方法是追根溯源去找到正在进行攻击的机器和攻击者。但是如果攻击者停止了攻击行为,则很难将其追踪到。唯一可行的方法是在其进行攻击的时候根据路由器的信息和攻击数据包的特征采用逐级回溯的方法来查找其攻击源头。

2. 协议欺骗攻击

协议欺骗有多种形式,常见的有 IP 欺骗、DNS 欺骗、源路由欺骗、ARP 欺骗等。

1) IP 欺骗

IP 欺骗就是一台主机设备冒充另一台主机的 IP 地址与其他设备通信,从而达到某种目的。攻击者选定目标主机 A 后,找到一台被目标主机信任的主机 B。首先使主机 B 丧失工作能力,并对主机 A 发出的 TCP 序列号进行采样猜测出它的数据序列号。然后伪装成主机 B,为了伪装成它,往往要先确保其不能接收到任何有效的网络数据。同时建立起与主机 A 基于地址验证的应用连接。如果成功,攻击者可以使用一种简单的命令放置一个系统后门,以进行非授权操作。

防范 IP 欺骗可以采用以下 3 种方法:

(1) 取消基于地址的信任策略。例如在 UNIX 系统中,不允许 r * 类远程调用命令的使用;删除 .rhosts 文件;清空/etc/hosts.equiv 文件。从而迫使所有用户使用其他远程通信手段,例如 Telnet、SSH、Skey 等。

(2) 进行包过滤。若用户的网络是通过路由器接入 Internet 的,就可以设置路由器来进行包过滤,可以过滤掉所有来自外部而希望与内部建立连接的请求。

(3) 使用加密方法。在通信时要求加密传输和认证。

2) DNS 欺骗

DNS 欺骗的基本原理是冒充域名服务器,将查询的 IP 地址设为攻击者的 IP 地址,用户上网就只能看到攻击者的主页,而非用户想要访问的网站主页。网络攻击者通常通过以下几种方法来进行 DNS 欺骗。

(1) 缓存感染。攻击者将数据放入一个没有设防的 DNS 服务器的缓存当中,这些缓存信息会在用户进行 DNS 访问时返回给用户,从而将用户引导到入侵者所设置的运行木马的 Web 服务器或邮件服务器上,攻击者就能从这些服务器上获取用户信息。

(2) DNS 信息劫持。入侵者通过监听客户端和 DNS 服务器的对话，猜测服务器响应给客户端的 DNS 查询 ID。每个 DNS 报文包括一个相关联的 16 位 ID 号，DNS 服务器根据这个 ID 号获取请求源位置。攻击者在 DNS 服务器响应之前将虚假的响应交给用户，从而欺骗客户端去访问恶意的网站。

(3) DNS 重定向。攻击者能够将 DNS 名称查询重定向到恶意 DNS 服务器，这样攻击者可以获得 DNS 服务器的写权限。

为了防范 DNS 欺骗，可以采取以下措施：

(1) 直接用 IP 访问重要的服务。

(2) 加密所有对外的数据流，但是该方法通常不容易实现。

3) 源路由欺骗

源路由欺骗是指通过指定路由，以假冒身份与其他主机进行合法通信或发送假报文，使受攻击主机出现错误动作。在通常情况下，信息包从起点到终点走过的路径是由位于该两点之间的路由器决定的，数据包只知道要去往的目的端，但并不知道具体的路线。源路由可使信息的发送者将此数据包要经过的路径写在数据包里，使数据包循着一个对方不可预料的路径到达目的主机。

例如，主机 A 享有主机 B 的某些特权，攻击者的主机 X 想冒充主机 A 从主机 B 获得某些服务。攻击者首先修改距离主机 X 最近的路由器，使得到达此路由器且包含目的地址主机 B 的数据包以主机 X 所在的网络为目的地，然后攻击者利用 IP 欺骗向主机 B 发送源路由（指定最近的路由器）数据包。当 B 回送数据包时，就传输到被更改过的路由器。这就使得攻击者可以假冒一个主机的名义通过一个特殊的路径来获得某些被保护数据。

以下几种防范措施可用来防止源路由欺骗：

(1) 最好的办法是配置好路由器，使其过滤从外部网进来的却声称是内部主机的报文。

(2) 对端路由器进行身份认证和路由信息的身份认证。

(3) 访问控制。对路由器的访问控制，需要进行密码的分级保护；基于 IP 地址的访问控制；基于用户的访问控制。

(4) 信息隐藏。对端通信时，不一定需要用真实身份进行通信。通过地址转换，可以做到隐藏网内地址、只以公共地址的方式访问外部网络。除了由内部网络首先发起的连接，网外用户不能通过地址转换直接访问网内资源。

(5) 在路由器上关闭源路由，例如使用命令 no ip source-route。

(6) 在路由器上提供攻击检测，可以防止一部分的攻击。

4) ARP 欺骗

在局域网中，通信前必须通过 ARP 协议将 IP 地址转换成第二层物理地址（即 MAC 地址）。而 ARP 欺骗就是通过伪造 IP 地址和 MAC 地址的对应关系实现 ARP 欺骗的攻击。

假设攻击者使用主机 A，与正在通信的主机 B、C 位于同一个交互式局域网中。A 伪装成 C 对 B 做 ARP 欺骗，即向 B 发送伪造的 ARP 应答包，应答包中的 IP 地址为 C 的 IP 地址而 MAC 地址为 A 的 MAC 地址。这个应答包会刷新 B 的 ARP 缓存，让 B 以为 A 就是 C。B 想要发送给 C 的数据实际上发送给了 A，这就达到了嗅探的目的。之后 A 还必须将数据转发给 C，以保证 B 和 C 间的通信不被中断。

可用来防止 ARP 欺骗的措施有以下几种：

(1) 在客户端使用 arp 命令绑定网关的真实 MAC 地址。

(2) 在交换机上进行端口与 MAC 地址的静态绑定。

(3) 在路由器上进行 IP 地址与 MAC 地址的静态绑定。

(4) 使用 ARP SERVER 按一定的时间间隔广播网段内所有主机的正确 IP-MAC 映射表。

3. 密码猜测攻击

密码是网络系统的第一道防线。当前的网络系统都是通过密码来验证用户身份、实施访问控制的。而密码猜测是一种出现概率很高的风险,因为它几乎不需要任何攻击工具,利用一个简单的暴力攻击程序和一个比较完善的字典,就可以猜测密码。一旦成功,攻击者进入了目标系统,他就能窃取、破坏和篡改被侵入方的信息,甚至完全控制被侵入方。因此,密码猜测攻击是攻击者实施网络攻击最基本、最重要、最有效的方法之一。

密码猜测攻击的主要方法有以下 5 种:

1) 网络嗅探

通过嗅探器在局域网内嗅探明文传输的密码字符串。避免此类攻击的对策是网络传输采用加密传输的方式进行。

2) 猜测攻击

密码猜测程序往往根据用户定义密码的习惯猜测用户密码。在详细了解用户的社会背景或私人资料之后,可以在短时间内完成猜测攻击。

3) 字典攻击

如果猜测攻击不成功,入侵者会继续扩大攻击范围,对所有英文单词进行尝试,程序将按序取出一个又一个单词,进行一次又一次尝试,直到成功。如果用户的密码不太长或是单词、短语,那么很快就会被破译出来。

4) 穷举法攻击

如果字典攻击仍然不能够成功,入侵者会采取穷举攻击。一般从长度为 1 的密码开始,按长度递增进行尝试攻击。由于用户偏爱简单易记的密码,穷举攻击的成功率很高。如果每千分之一秒检查一个密码,那么 86%的密码可以在一周内破译出来。

5) 直接破解系统密码文件

当所有的攻击都不能成功时,攻击者会寻找目标主机的安全漏洞和薄弱环节,窃取存放系统密码的文件,破译加密的密码之后冒充合法用户访问这台主机。

防范密码猜测攻击的方法:

(1) 好密码是防范密码猜测攻击的最基本、最有效的方法。最好采用字母、数字、还有标点符号、特殊字符的组合,同时有大小写字母,长度最好达到 8 个以上,最好容易记忆,绝对不要使用自己或亲友的生日、手机号码等易于被他人获知的信息作密码。

(2) 要注意保护密码安全。不要将密码记在纸上或存储于计算机文件中;最好不要将密码告知他人;不要在不同的系统中使用相同的密码;在输入密码时应确保无人在身边窥视;在公共上网场所最好先确认系统是否安全;定期更改密码等。这样才能使自己遭受密码攻击的风险降到最低。

4. 木马程序攻击

特洛伊木马是一个程序,它驻留在目标计算机里,可以随计算机自动启动并在某一端口

进行侦听,在对接收的数据识别后,对目标计算机执行特定的操作。木马的实质是一个通过端口进行通信的客户机/服务器程序。对于特洛伊木马,被控制端就成为一台服务器,控制端则是一台客户机。

可采用以下几种方法来预防此类极为常见的攻击手段:

1) 端口扫描

端口扫描是检查远程机器有无木马的最好方法,扫描程序尝试连接某个端口,如果成功则说明端口开放;如果失败或超过某个特定的时间无响应,则说明端口关闭。但对于驱动程序/动态链接木马,扫描端口是不起作用的。

2) 查看连接

查看连接和端口扫描的原理基本相同,不过是在本地机上通过 netstat -a 查看所有的 TCP/UDP 连接,查看连接要比端口扫描快,但同样是无法查出驱动程序/动态链接木马,而且仅仅能在本地使用。

3) 检查注册表

木马可以通过注册表启动,那么同样可以通过检查注册表来发现木马在注册表里留下的痕迹。

4) 查找文件

查找木马特定的文件也是一个常用的方法,木马的一个特征文件是 kernl32. exe,另一个是 sysexlpr. exe,只要删除了这两个文件,木马就已经不起作用了。

5. Web 欺骗攻击

Web 欺骗是一种电子信息欺骗,错误的 Web 拥有相同的网页和链接,看起来十分逼真。攻击者通过控制错误的 Web 站点,截获受害者浏览器和 Web 之间的所有网络信息。攻击者可以观察或修改从受攻击者到 Web 服务器的任何信息,也控制着从 Web 服务器到受害者的返回数据,从而有机会实施监视和破坏。

Web 欺骗能够成功的关键是在受害者与其他 Web 服务器中间建立起攻击者的 Web 服务器,这种攻击种类在安全问题中称为“来自中间的攻击”。为了建立起这样的中间 Web 服务器,攻击者往往会对 URL 进行改写,这样它们指向了攻击者的 Web 服务器而不是真正的 Web 服务器。

Web 欺骗是当今 Internet 上具有相当危险性而不易被察觉的欺骗手法,可以采取的一些防范方法有:

(1) 禁止浏览器中的 JavaScript 功能,从而使得各类改写信息暴露出来。

(2) 确保浏览器的链接状态是可见的,它将给你提供当前位置的各类信息。

(3) 时刻注意所单击的 URL 链接会在位置状态行中得到正确的显示。

6. 邮件炸弹

邮件炸弹指的是邮件发送者利用特殊的电子邮件软件,在很短的时间内连续不断地将邮件邮寄给同一个收信人,这些数以千万计的电子邮件的总容量就会超过电子邮箱的总容量,以至造成邮箱超负荷而崩溃。

这种攻击手段不仅会干扰用户电子邮件系统的正常使用,还会大量消耗网络资源,往往导致网络阻塞,使大量的用户不能正常工作,甚至还会影响到邮件系统所在的服务器系统的

安全，造成整个网络系统全部瘫痪。

防范邮件炸弹的方法有以下几种：

(1) 不要将个人邮箱地址到处传播。

(2) 使用 Outlook 或 Foxmail 等 POP 收信工具收取 E-mail。

(3) 在收信时，一旦看见邮箱列表的数量超过平时正常邮件的数量的若干倍，应当马上停止下载邮件，然后从服务器删除炸弹邮件。

(4) 谨慎使用自动回信功能。炸弹邮件的发件人有可能是假地址，甚至可能填的与收件人地址相同。因此，邮件收取的回复功能对发送炸弹邮件的攻击者没有作用。

(5) 采用过滤功能。在接收任何电子邮件之前预先检查发件人的资料，如果觉得有可疑之处，可以将之删除，不让它进入用户的邮件系统，但这种做法有时会误删除一些有用的邮件。用户可以在邮件软件中启用过滤功能，自动将超过信箱容量的大邮件删除。

7. 缓冲区溢出攻击

缓冲区溢出是指计算机程序向缓冲区内填充的数据位数超过了缓冲区本身的容量，溢出的数据覆盖在合法数据上。理想情况是程序检查数据长度并且不允许输入超过缓冲区长度的字符串，但是绝大多数程序都会假设数据长度总是与所分配的存储空间相匹配，这就为缓冲区溢出埋下了隐患。操作系统所使用的缓冲区又被称为堆栈，在各个操作进程之间，指令被临时存储在堆栈当中，堆栈也会出现缓冲区溢出。

当一个超长的数据进入到缓冲区时，超出部分都会被写入其他缓冲区，然而其他缓冲区内存放的可能是数据、下一条指令的指针或者是其他程序的输出内容，这些内容都会被覆盖或者破坏掉。由此可见小部分数据或者一套指令溢出就可能导致一个程序或者操作系统崩溃。

为了防范缓冲区溢出攻击，可以采用如下措施：

(1) 编写正确的代码。用 grep 来搜索源代码中容易产生漏洞的库的调用，例如 strcpy 和 sprintf 这两个函数都没有检查输入参数的长度，应尽量使用 strncpy 和 memcpy 等含有处理长度参数的函数。

(2) 非执行的缓冲区。通过使被攻击程序的数据段地址空间不可执行，从而使攻击者不可能执行植入在被攻击程序输入缓冲区的代码。

(3) 数据边界检查。应当检查所有对数据的读写操作以确保在正确的范围内操作数组。

(4) 程序指针完整性检查。与数组边界检查相比，这种方法不能解决所有的缓冲区溢出问题。但是该方法在性能上有很大的优势，而且兼容性也很好。

8. Windows 系统漏洞攻击

Windows 操作系统是现在使用较多的操作系统，而其中不可避免的一些漏洞也一直都是各种攻击者的入侵目标。为了防止 Windows 系统漏洞攻击，可以采取如下防范措施：

(1) 管理好 Administrator 账户。给 Administrator 账户设置强壮的安全密码。

(2) 关闭不用的端口和共享服务。将无用的端口服务关闭掉，避免被攻击者利用。

(3) 备份进程和初始进程列表。不管是病毒还是木马，只要侵入到用户系统，都会在其系统内加载木马启动项，以便机器重启后会自动调用到木马。给系统做一个备份快照，当有

木马侵入时，只要将之前备份的数据和当前数据相对比，就会很容易地发现入侵的可疑木马。

(4) 利用工具。为了能做好系统的全面检测工作，可以借助安全工具为系统做详细的检查，从而更有效地保护系统。

9. SQL 注入

SQL 注入攻击技术是利用程序员对用户输入数据的合法性检测不严或不检测的特点，故意从客户端提交特殊的不合法的 SQL 代码，让服务器报错，通过报错信息来收集程序及服务器的信息，从而获取想得到的资料。通过反复尝试，攻击者可以从出错信息中获得数据库类型、用户账户、管理员的账户和密码、数据库的库名和数据表名，然后就可以入侵网站了。因此攻击者必须要先确认一些服务器产生的错误提示类型。

防范 SQL 注入攻击的方法有如下几种：

(1) 在服务器端正式处理之前对提交数据的合法性进行检查。

(2) 封装客户端提交信息。

(3) 替换或删除敏感字符/字符串。

(4) 屏蔽出错信息。

10. UNIX 系统攻击

攻击者入侵一台 UNIX 主机一般可以分为 6 个阶段：

(1) 信息的收集。通过各种方式获得目标的信息，常见的方式有端口扫描、用户名收集、密码猜测、远程溢出攻击等。

(2) 取得普通用户的权限。攻击者即使只获得了很小的权限，仍可以利用各种漏洞来提升其权限，从而最终获得 root 权限。

(3) 远程登录。

(4) 取得中级用户的权限。

(5) 留下后门。

(6) 清除日志。

针对 UNIX 系统攻击的防范措施有以下几点：

(1) 定期检查属性为-rwsr-sr--的文件。这种文件一旦被 user 执行，就会具有和该文件创造者一样的权限。如果文件的创造者为 root，那攻击者马上可以获得相同的权限。攻击者一般都将 shell 改成此类属性然后隐藏起来，便于下次利用。

(2) 使用 getsniff 和 rookit detecor 等工具查找系统中是否有嗅探器和 rookit 黑客工具包。

(3) 如果不需要 FTP 服务，最好关掉。

(4) 用户要经常了解系统是否又有新漏洞出现，及时打上补丁。

12.3　入侵检测技术

12.3.1　入侵检测系统的概念

入侵检测技术是对防火墙技术的进一步补充，入侵检测系统(Intrusion Detection

System，IDS）对计算机网络和计算机系统的关键节点的信息进行收集和分析，检测其中是否有违反安全策略的事件发生或攻击迹象，并通知系统安全管理员。通常将用于入侵检测的软件和硬件统称为入侵检测系统。入侵检测系统能在入侵攻击发生危害之前，检测并通过报警与防护系统阻断攻击；在入侵攻击发生之时，减少其造成的损失；在入侵攻击发生之后，收集入侵的相关信息作为防范系统的知识，以增强系统的防范能力。

入侵检测系统的分类方式有：

（1）根据其检测的对象是主机还是网络分为基于主机的入侵检测系统和基于网络的入侵检测系统。

（2）根据检测系统对入侵行为的响应方式分为主动检测系统和被动检测系统。

（3）根据工作方式分为在线检测系统和离线检测系统。

12.3.2 入侵检测系统的功能

单纯的防火墙存在许多不足和弱点，例如，不能防范如 TCP/IP 协议等本身存在的漏洞；无法解决安全后门问题；不能阻止网络内部攻击；不能提供实时入侵检测能力等。IDS 系统弥补了防火墙的不足，其目的是帮助系统对付内部攻击、外部攻击以及误操作的实时保护，在网络系统受到危害之前拦截和响应入侵，例如记录证据、跟踪入侵或断开网络连接等。IDS 系统提高了系统管理员的安全管理能力，有助于提高信息安全基础结构的完整性。同时，IDS 系统以旁路侦听的方式收集和分析信息，在不影响网络性能的情况下对网络进行检测。图 12-4 为入侵检测系统示意图。

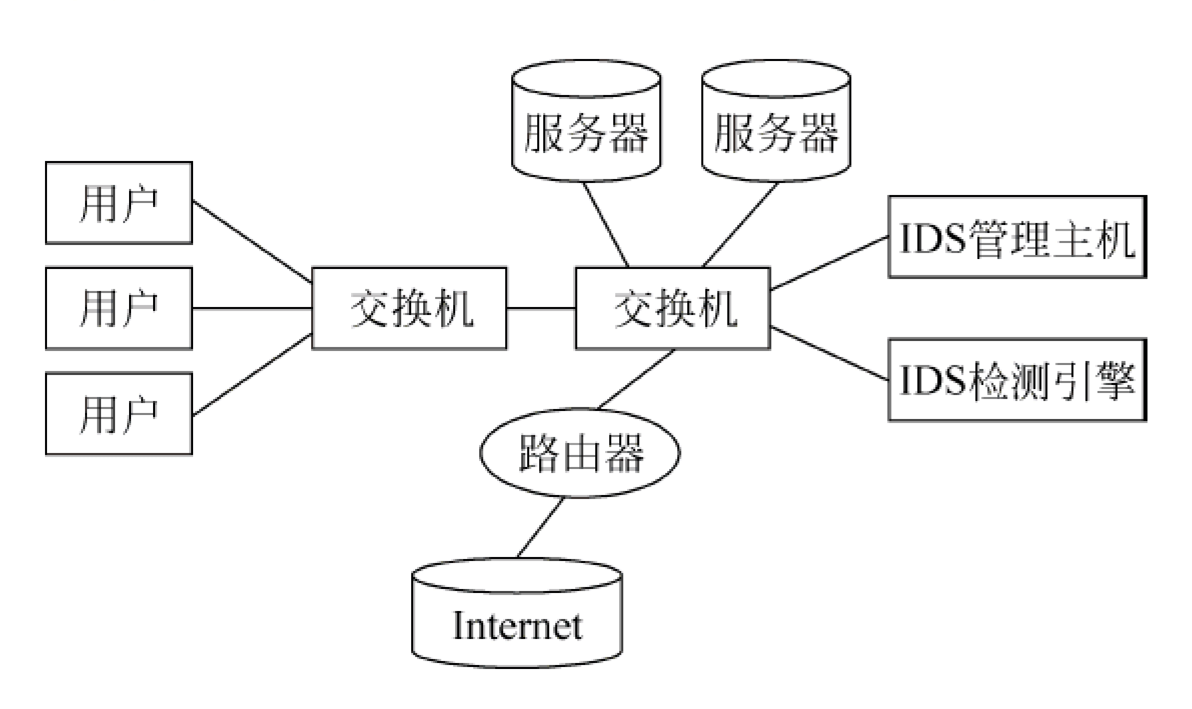


图 12-4 入侵检测系统

不同于防火墙，IDS 是一个监听设备，没有跨接在任何链路上，无须网络流量流经它便可以工作。因此，对 IDS 的部署唯一的要求是：IDS 应当挂接在所有所关注流量都必须流经的链路上。“所关注流量”指的是来自高危网络区域的访问流量和需要进行统计、监视的网络报文。在如今的网络拓扑中，已经很难找到以前的 HUB 式的共享介质冲突域的网络，绝大部分的网络区域都已经全面升级到交换式的网络结构。因此，IDS 在交换式网络中的位置一般选择在尽可能靠近攻击源或尽可能靠近受保护资源的地方。这些位置通常是服务器区域的交换机、Internet 接入路由器之后的第一台交换机、重点保护网段的局域网交换机等。

IDS 系统功能主要有：

（1）识别常见入侵与攻击。检测并分析系统和用户的活动，查找非法用户和合法用户

的越权操作。IDS 系统通过分析各种攻击特征,可以全面快速地识别探测攻击、拒绝服务攻击、缓冲区溢出攻击、电子邮件攻击、浏览器攻击等各种常用攻击手段,做出相应防御的同时向管理员发出警告。

(2) 监控网络异常通信。IDS 系统会对网络中不正常的通信连接做出反应,保证网络通信的合法性;任何不符合网络安全策略的网络数据都会被 IDS 监测到并发出警告。

(3) 鉴别对系统漏洞及后门的利用。IDS 系统一般带有系统漏洞及后门的详细信息,通过对网络数据包连接的方式、连接端口以及连接中特定的内容等特征分析,可以有效地发现网络通信中针对系统漏洞进行的非法行为。

(4) 完善网络安全管理。IDS 通过对攻击或入侵的检测和反应,可以有效地发现和防止大部分的网络入侵或攻击行为,给网络安全管理提供了一个集中、方便和有效的工具。使用 IDS 系统的检测、统计分析、报表功能可以进一步完善网络管理。

图 12-5 所示为入侵检测系统(IDS)的功能示意图。

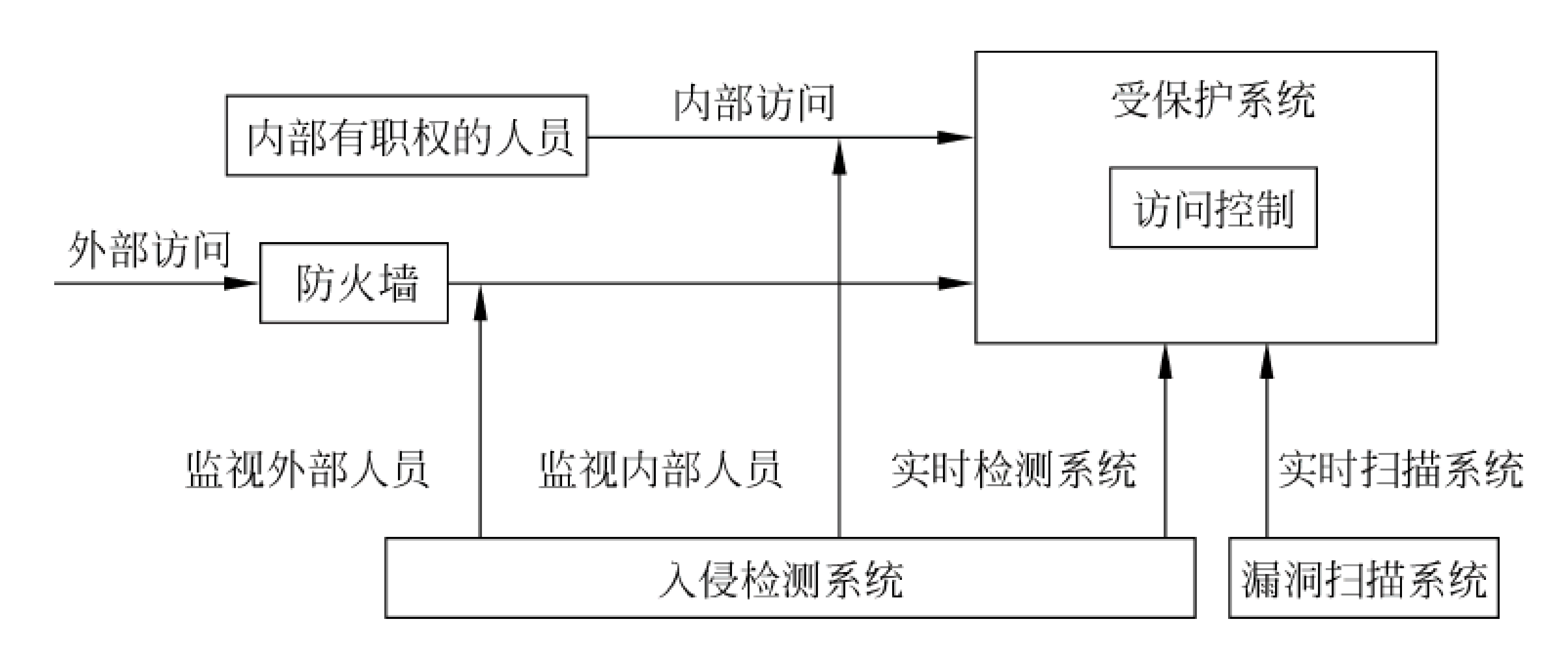


图 12-5　入侵检测系统(IDS)的功能示意图

12.3.3　入侵检测系统的原理、结构和流程

1. 入侵检测系统的原理

入侵检测系统实际上可以看成一个窥探设备,它不跨接多个物理网段,无须转发任何流量,只需要在网络上收集它所关心的报文,并从其中提取相应的流量统计特征值,并利用内置的入侵知识库与这些流量特征进行智能分析比较匹配,判断出其中的攻击行为,并对其进行报警或相应的反击。

2. 入侵检测系统的结构

IDS 大体上都可分为 4 个部分:传感器模块、数据处理分析模块、管理与控制模块和数据库。

1) 传感器模块

传感器可以是安装在保护主机的软件,也可以是通过网线连接到要保护的网络里的硬件。传感器的作用是负责采集数据(例如网络数据包、系统日志、网络行为数据等),并负责对数据进行预处理,并将预处理后的数据传输给数据处理和分析模块。对于大型网络或者分布式入侵检测系统,传感器可能有多个,需保护的区域均要安装数据采集传感器。

2) 数据处理分析模块

数据处理分析模块负责对传感器获取的数据做进一步处理,例如特征检测、统计分析、

智能推理等。主要目的是发现入侵、攻击或违反安全策略的行为或事件，并将处理结果传输给管理和控制模块。

3）管理和控制模块

管理和控制模块主要负责系统的总体控制和检测分析结果的进一步处理，包括日记记录、实时报警或实施有限度的反击等。管理和控制模块在 IDS 中处于非常重要的地位。

4）数据库

数据库主要用于存储入侵、攻击或违反安全策略的行为模式、模板或特征数据等。

3. 入侵检测系统的工作流程

IDS 系统的工作流程大体可分为 3 个步骤：信息收集、信息分析和结果处理。

1）信息收集

收集的信息内容包括：网络流量的内容、日志、用户连接活动的状态和行为以及历史数据等。

2）信息分析

信息分析的一般方法有 3 种：模式匹配、统计分析和完整性分析。

(1) 模式匹配。模式匹配是将收集到的信息与已知的网络入侵和系统误用模式数据库进行比较，从而发现违背安全策略的行为。该方法的一大优点是只需收集相关的数据集合，显著减少系统负担，且技术较为成熟。与病毒防火墙采用的方法类似，模式匹配检测准确率和效率都相当高，但它的弱点是需要不断地升级以实现对不断出现的新的攻击方法的检测。

(2) 统计分析。统计分析方法首先对信息对象创建一个统计描述，测量属性的平均值将被用来与网络、系统的行为进行比较，任何观察值在正常偏差之外时都认为有入侵发生。其优点是可检测到位置的入侵和更为复杂的入侵，缺点是误报、漏报率高，且不适应合法用户正常行为的突然改变。具体的统计分析方法有基于专家系统的、基于模型推理的和基于神经网络的方法等。

(3) 完整性分析。该方法主要关注某个文件或对象是否被更改，包括文件和目录的内容及数据，在发现被更改的、被破坏的应用程序方面特别有效。完整性分析利用强有力的加密机制(称为消息摘要函数例如 MD5)，能识别极其微小的变化。其优点是不管模式匹配方法和统计分析方法能否发现入侵，只要是成功的攻击导致了文件或其他对象的任何改变，它都能够发现。缺点是一般以批处理方式实现，不用于实时响应。这种方式主要应用于基于主机的 IDS 系统。

3）结果处理

IDS 的根本任务是要对入侵行为做出适当的反应，这些反应包括详细日志记录、实时报警和有限度地反击攻击源。

12.3.4 入侵检测技术分类与检测模型

入侵检测技术通过对入侵行为的过程与特征的研究，使安全系统对入侵事件和入侵过程能做出实时响应，主要包括特征检测、异常检测和协议分析。

1. 基于标识的特征检测技术

基于标识的特征检测技术又称为误用检测，首先定义违背安全策略事件的特征，再根据

这些特征来检测主体活动。如果主体活动具有这些特征，就可以认为该主体活动是入侵行为。特征检测模型如图 12-6 所示。

该模型的优点是算法简单、系统开销小、准确率高、效率高。缺点有：只能检测出已知攻击，新类型的攻击会对系统造成很大的威胁；模式库的建立和维护难，知识依赖于硬件平台、操作系统、系统中运行的应用程序。

2. 基于异常情况的检测技术

异常检测的假设是入侵者活动异常于正常主体的活动，建立正常活动的活动档案，当前主体的活动违反档案的统计规律时，认为该活动可能是入侵行为。异常检测模型如图 12-7 所示。

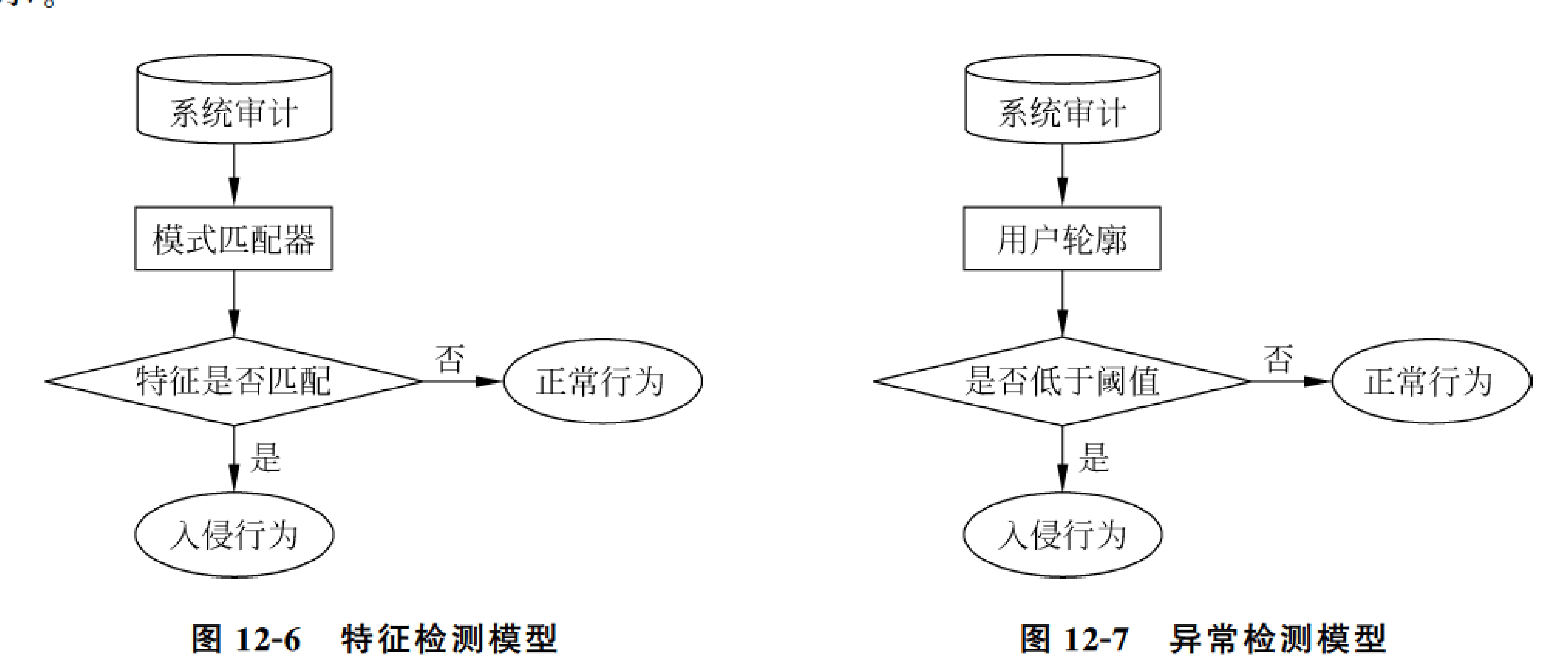


图 12-6　特征检测模型　　**图 12-7　异常检测模型**

其中，用户轮廓(Profile)通常定义为各种行为参数及其阈值的集合，用于描述正常行为范围。该模型的优点是：可以检测到未知的入侵；可以检测冒用他人账户的行为；具有自适应、自学习功能；无须系统先验知识。缺点是：漏报、误报率高；统计算法的计算量庞大，效率很低；统计点的选取和参考库的建立比较困难。

3. 协议分析

协议分析技术是新一代 IDS 系统探测攻击手法的主要技术，也是目前比较流行的检测技术。它利用网络协议的高度规则性并结合高速数据包捕捉、协议分析和命令解析，以快速探测攻击的存在。

图 12-8 所示为通用入侵检测模型。

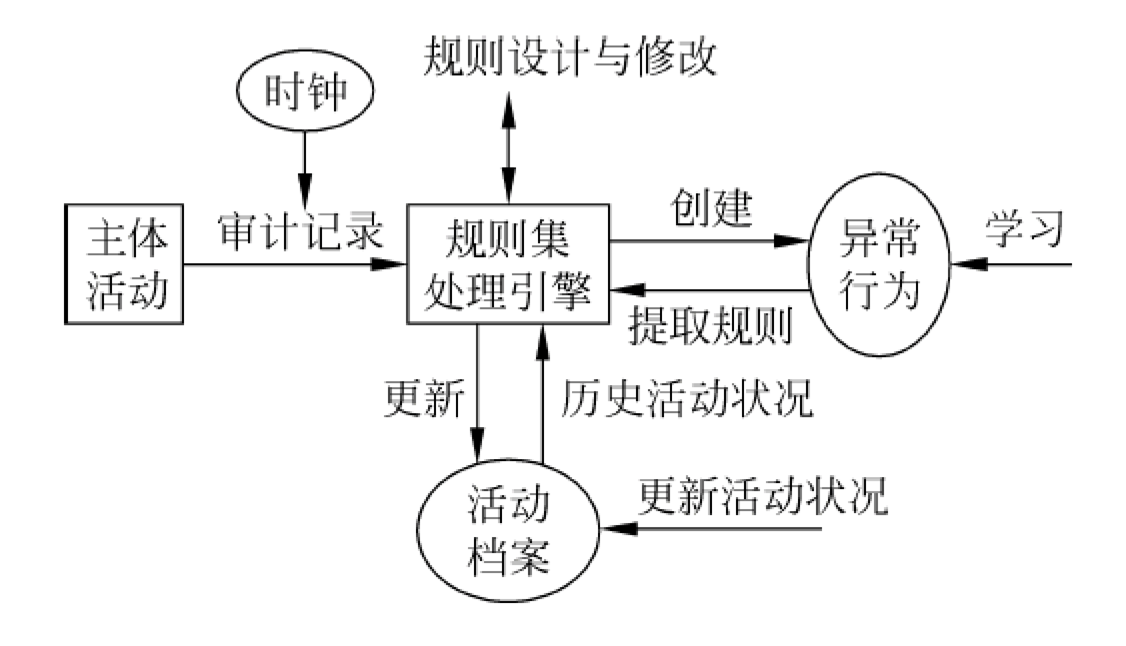


图 12-8　通用入侵检测模型

12.3.5 入侵检测系统的设置

入侵检测系统的设置主要分5个步骤：

(1) 确定入侵检测需求。

(2) 设计入侵检测系统在网络中的拓扑位置。

(3) 配置入侵检测系统。

(4) 入侵检测系统磨合。

(5) 入侵检测系统的使用及自调节。

入侵检测系统设置流程如图12-9所示。

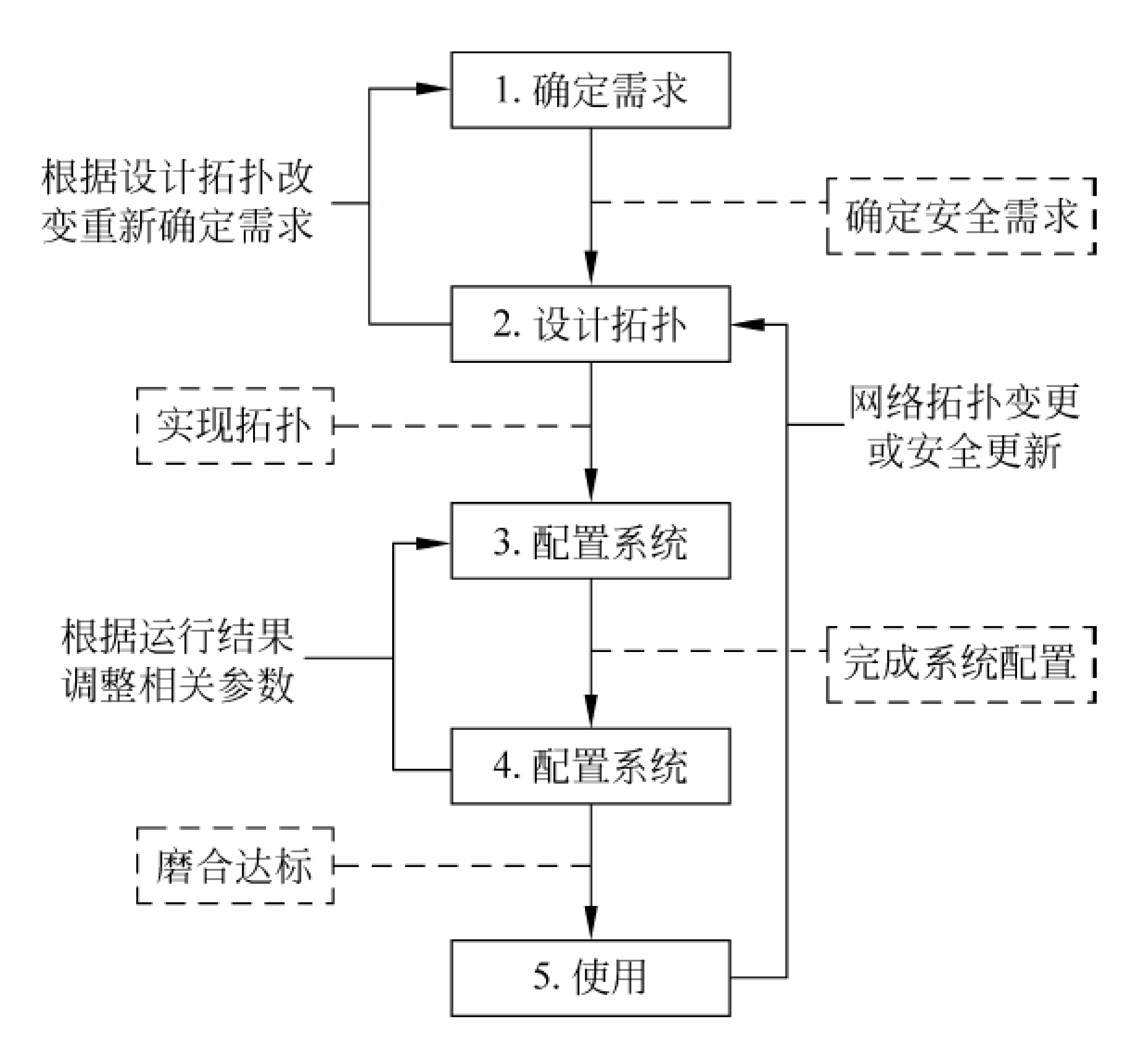


图12-9 入侵检测系统设置流程图

12.3.6 入侵检测技术的未来发展

随着入侵或攻击行为的综合化、复杂化，入侵主体对象的间接化，入侵或攻击规模、范围的扩大，入侵或攻击技术的分布化，以及攻击对象的不断转移，入侵检测技术将随着入侵或攻击的变化而不断演化和发展，如下几个方面需要加以考虑。

1. 大范围的分布式入侵检测

针对分布式网络攻击的检测方法，使用分布式的方法来检测分布式攻击，其中的关键技术是检测信息的协同处理与入侵攻击的全局信息获取。

2. 智能化入侵检测

使用智能化的方法与手段来进行入侵检测，神经网络、遗传算法、模糊技术、免疫原理等方法可能用于入侵特征的辨识与泛化，智能代理技术可能广泛应用于入侵检测技术。

3. 系统层的安全保障体系

将其他安全技术融入到入侵检测系统中，或者入侵检测系统与其他网络安全设备进行互动、互相协作，构成较为全面的安全保障体系。

12.4　取证技术

12.4.1　计算机取证的基本概念

随着信息技术的不断发展，与计算机相关的法庭案例也不断出现，一种新的存在于计算机及相关外围设备(包括网络介质)中的电子证据组件成为新的诉讼证据之一。商业机密信息的窃取和破坏、计算机欺诈、对政府或金融网站的破坏等计算机犯罪案例时常发生，这些案例的取证工作需要提取存在于计算机系统中的数据，甚至需要从已被删除、加密或破坏的文件中重新获得信息。电子证据本身和取证过程都有着许多有别于传统物证和取证的特点，这对司法和计算机科学领域提出了新的挑战。作为计算机领域和法学领域的一门交叉学科，计算机取证正逐渐成为人们研究与关注的焦点。

1. 计算机取证的定义

计算机在相关的犯罪案例中可以扮演黑客入侵的目标、作案的工具和犯罪信息的存储器这 3 种角色。无论作为哪种角色，计算机(连同其外设)中会留下大量与犯罪有关的数据。对计算机犯罪的证据进行获取、保存、分析和出示，计算机取证实际上是一个详细扫描计算机系统以及重建入侵事件的过程。

计算机取证包括物理证据获取和信息发现两个阶段。物理证据获取是指调查人员来到计算机犯罪或入侵的现场，寻找并扣留相关的计算机硬件。信息发现是指从原始数据(包括文件、日志等)中寻找可以用来证明或反驳的证据。与其他证据一样，电子证据必须是真实、可靠、完整和符合法律规定的。

综合以上说法，计算机取证是指对能够为法庭接受的、足够可靠和有说服力的、存在于计算机和相关外设中的电子证据的确认、保护、提取和归档的过程，它能推动或促进犯罪事件的重构，或者帮助预见有害的未经授权的行为。从动态的观点来看，计算机取证可以归结为以下几点：在犯罪进行过程中或之后搜集证据；重构犯罪行为；为起诉提供证据。

2. 电子证据的概念

电子证据是计算机取证技术的核心，它与传统证据的不同之处在于它以电子介质为媒介。电子证据往往以多种形式存在：电子文件、图像文件、视频文件、隐藏文件、电子邮件、光盘、网页和域名等。而且其涉及的领域很广，例如证明著作权侵权、不正当竞争以及经济诈骗等。随着网络技术的快速发展，还出现了许多除电子文件和邮件之外的新型电子证据，例如 CRM(客户关系管理系统)，可以管理和记录客户在网上的一切活动和特征(例如浏览内容、停留时间和收发信息等)。在法律上如何准确定位这些新型的电子记录以及上文提及的电子邮件、多媒体软件、网页的地位和证明效力，将是清晰解决各种网络纠纷的前提。

目前，国内法学界多数学者将电子证据定义为：在计算机或计算机系统运行过程中产生的以其记录的内容来证明案件事实的电磁记录物。

3. 电子证据的特点

电子证据与其他种类的证据相比具有如下特点：

(1) 表现形式的多样性。电子证据不仅可以用文字、图像和声音等多种方式存储，还可

以以多媒体形式存在。将多种表现形式融为一体是电子证据所特有的特点。

(2) 存储介质的电子性。电子证据依据计算机技术产生、转化为电子信息存储在特定的电子介质上,例如硬盘或光盘。证据的产生和重现必须依赖于这些特定的电子介质,这也正是电子证据的弱点。如果改变了电子介质,就会改变电子证据,给证据的认定造成困难。

(3) 准确性。电子信息严格按照运行于计算机上的各种软件和技术标准产生和运行,不会受到主观因素的影响。在没有人为的蓄意修改或破坏的情况下,具有非常高的准确度。

(4) 脆弱性。电子证据存储在电子介质上,很容易被修改,甚至操作人员误操作或系统故障时都有可能对电子证据造成难以修复的损坏。

(5) 数据的挥发性。数据的挥发性是指在计算机系统中,有些紧急事件的数据必须在一定的时间内获得才有效,即经过一段时间数据可能就无法得到或失效了。因此,在收集电子证据时必须充分考虑到数据的挥发性,在数据的有效期内及时收集数据。

电子证据和传统证据相比,具有以下优点:

(1) 可以被精确地复制。即可以只需对复件进行检查分析,避免原件受到损坏,同时还可以避免原始证据被恶意销毁。

(2) 用适当的软件工具和原件对比,很容易鉴别出当前的电子证据是否有改变。

4. 常见的电子证据

随着各类电子设备的广泛应用,电子证据几乎无所不在。下面介绍几种常见电子设备中潜在的电子证据。

1) 计算机系统

计算机系统的硬盘及其他存储介质就是电子证据最为常见的载体。在移动存储器(外置硬盘、移动硬盘、U 盘、光盘等)中需要检查的应用数据包括:用户自建的文档(地址簿、视/音频文件、图片影像文件、Internet 书签/收藏、数据库文件和文本文件等);用户保护文档(压缩文件、改名文件、加密文件、密码保护文件和隐藏文件等);计算机创建的文档(备份文档、日志文件、配置文件、Cookies、交换文件、系统文件、隐藏文件、历史文件和临时文件等);其他的数据区上可能存在的数据证据(硬盘上的坏簇、其他分区、lack 空间、计算机系统时间和密码、被删除的文件、软件注册信息、自由空间、隐藏分区、系统数据区、丢失簇和未分配空间)。此外,计算机附加控制设备(例如智能卡和加密狗等)具有控制计算机输入/输出或加密功能,这些设备可能含有用户的身份和权限等重要信息。

2) 联网设备

联网设备包括各类调制解调器、网卡、路由器、集线器、交换机等。从设备中也可以获得配置文件等一些重要信息。

3) 手持电子设备

个人数字助理(PDA)、电子记事本等设备中可能包含有地址簿、文本信息等。数码相机、可视电话等设备可能存储有影像、视频等信息。

12.4.2 计算机取证方法分类

根据取得证据的用途可以分为两类不同性质的取证:来源取证和事实取证。

1. 来源取证

来源取证是指取证的目的主要是确定犯罪嫌疑人或证据的来源。这类取证主要有 IP

地址取证、MAC 地址取证、电子邮件取证、软件账户取证等。

(1) IP 地址取证。利用在 Internet 中每一台联网的机器在某一时刻都有唯一的全局 IP 地址,根据在案发现场找到的 IP 地址信息,进一步确定犯罪嫌疑人的机器,由犯罪嫌疑人的机器再寻找案件相关人的方法。

(2) MAC 地址取证。主要在一些局域网或动态分配 IP 地址网络中,根据物理地址与逻辑地址的关系找到物理地址。由于 MAC 地址与特定计算机设备中的网卡存在一定的对应关系,可以用来确定来源。

(3) 电子邮件取证。根据电子邮件头部信息找到发送电子邮件的机器,并根据已锁定的机器找到犯罪嫌疑人。

(4) 软件账户取证。特定软件如果其某个账户与特定人存在一一对应关系时也可以用来证明案件的来源。

2. 事实取证

事实取证是指取证目的不是为了查明犯罪嫌疑人,而是取得与证明案件相关事实的证据。常见的取证方法有文件内容调查、使用痕迹调查、软件功能分析、软件相似性分析、日志文件分析、网络状态分析、网络数据包分析等。

(1) 文件内容调查。在存储设备中取得文档、图片、音频、视频、动画、网页、电子邮件等相关文件的内容。包括这些文件被删除以后、文件系统被格式化以后或者数据恢复以后的文件内容。

(2) 使用痕迹调查。包括 Windows 运行的痕迹(包括运行栏历史记录、搜索栏历史记录、打开/保存文件记录、临时文件夹、最近访问的文件等使用文件与程序调查)、上网记录的调查(缓存、历史记录、自动完成记录、浏览器地址栏下拉网址,Cookies、index、dat 文件等等)、Office、Real Player 和 Media Player 的播放列表及其他应用软件使用历史记录。

(3) 软件功能分析。该方法主要针对特定软件和程序的性质和功能进行分析,常见的是对恶意代码的分析,确定其破坏性、传染性等特征。通常在破坏计算机信息系统、入侵计算机信息系统、传播计算机病毒行为分析中使用此类取证方法。

(4) 软件相似性分析。比较两个软件,找出二者之间是否存在实质性相似的证据,用来解决软件知识产权的纠纷。

(5) 日志文件分析。通过系统日志、数据库日志、网络日志、应用程序日志等进行分析发现系统是否存在入侵行为或者其他访问行为的证据。

(6) 网络状态分析。取得特定时刻计算机联网状态,例如网络中哪些机器与本机相连、本机的网络配置、开启了哪些服务、哪些用户登录到本机等信息。

(7) 网络数据包分析。通过分析网络中传输的数据包发现相关证据的过程。网络数据包分析主要发生在实时取证中,是一种综合的取证方法。有时候网络数据包分析也称为“网络侦听”。在对网络犯罪实时侦查或“诱惑性”侦查时,往往采取网络侦听的方法发现犯罪嫌疑人的犯罪活动,掌握犯罪的线索,为抓获犯罪嫌疑人提供支持。

12.4.3　计算机取证的原则、一般步骤和取证模型

1. 取证的基本原则

计算机取证要遵循以下基本原则:

(1) 早搜集证据,并保证其没有受到任何破坏,也不会被取证程序本身所破坏。

(2) 必须保证取证过程中计算机病毒不会被引入目标计算机。

(3) 必须保证证据连续性,要求证据从最初的获取状态到作为证据时的状态之间没有变化,要求说明证据的取证复制是完全的,复制证据的过程是可靠并可复验的,同时所有的介质都是安全的。

(4) 整个检查、取证过程必须是受到监督的。

(5) 必须保证提取出来的可能有用的证据不会受到机械或电磁损害。

(6) 被取证对象如果必须运行某些商务程序,要确保该程序的运行只能影响有限时间。

(7) 尊重不小心获取的任何关于代理人的私人信息,不能将这些信息泄露出去。

2. 计算机取证的阶段及步骤

1) 计算机取证的阶段

计算机取证的过程一般可划分为3个阶段:获取、分析和陈述。

(1) 获取阶段:获取阶段保存计算机系统的状态,以供日后分析。这一阶段的任务是保存所有电子数据,至少要复制硬盘上所有已分配和未分配的数据,即通常所说的映像,要保证数据的完整性。

(2) 分析阶段:分析已获得的数据,并分析这些数据确定证据的类型。类型分为:判断有罪的证据;辨明无罪的证据;已篡改了的证据。

(3) 陈述阶段:陈述阶段将给出调查所得结论及相应的证据,这一阶段应依据政策法规行事,对不同的机构采取不同的方法。

2) 计算机取证的步骤

根据上述3个阶段的特点,计算机取证一般按照以下几个步骤进行。

(1) 在取证检查中,保护目标计算机系统,避免发生任何的改变、伤害、数据破坏或病毒感染。

(2) 搜索目标系统中的所有文件。包括现存的正常文件、已经被删除但仍存于磁盘上(即还没有被新文件覆盖)的文件、隐藏文件、受到密码保护的文件和加密文件。

(3) 全部(或尽可能)恢复所发现的已删除文件。

(4) 最大限度地展示操作系统或应用程序使用的隐藏文件、临时文件和交换文件内容。

(5) 在法律允许的前提下,访问被保护或加密文件的内容。

(6) 分析磁盘特殊区域以发现有价值的数据。该特殊区域包括未分配的磁盘空间和文件的空白空间。

(7) 打印对目标计算机系统的全面分析结果,以及所有可能有用的文件和被挖掘出来的文件数据和清单。在此基础上给出分析结论,包括系统的整体情况,已发现的文件结构、被挖掘出来的数据和作者的信息,对信息的任何隐藏、删除、保护和加密企图,以及在调查中发现的其他的相关信息。

(8) 给出必要的专家证明和/或在法庭上的证词。整个计算机取证过程应当有计算机取证专家的指导和监督。

3. 计算机取证模型

对网络攻击行为模式的研究有助于对网络入侵取证途径的研究。无论攻击者的技术水

平如何，网络攻击通常遵循同一个行为模式，一般都要经过寻找攻击目标、入侵、破坏和掩盖入侵足迹等几个攻击阶段。图 12-10 所示为传统取证周期的状态转移模型。在传统模型中，在犯罪发生之后或犯罪被发现之后才开始进行取证。前文提到的计算机取证原则及步骤都是基于这一模型。而在图 12-11 中，周期从取证行为开始，即允许在犯罪发生前开始搜集证据。大量的事后取证行为是由对已搜集证据的分析组成。对网络入侵行为的取证往往应采取后一种模型，否则在入侵者消除入侵足迹后再进行证据获取就为时已晚。这种模型的实现往往需要将计算机取证工具和 IDS、蜜罐等网络安全工具相结合。

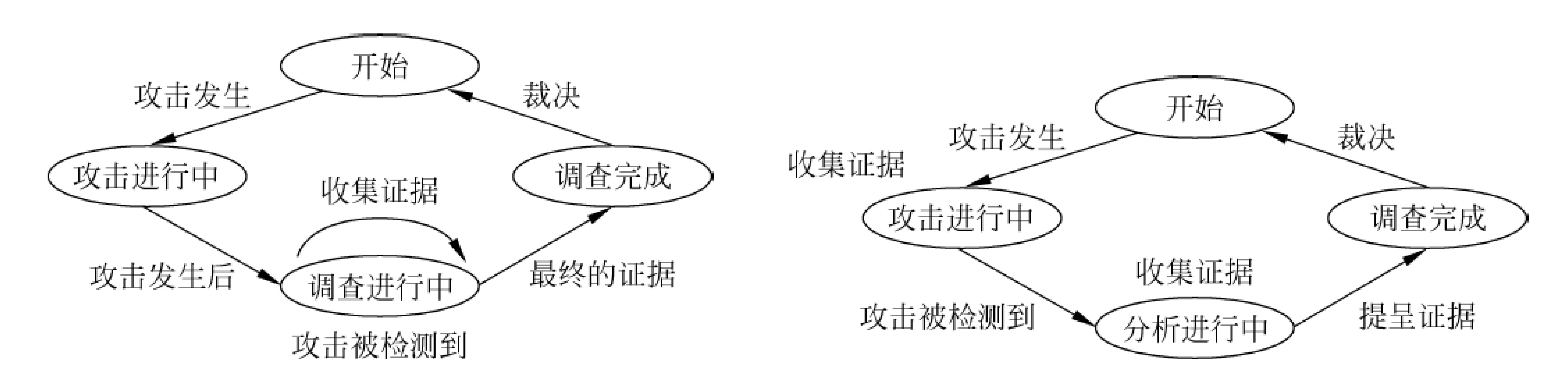


图 12-10　传统取证周期状态转移模型　　**图 12-11　计算机取证周期的状态转移模型**

12.4.4　计算机取证的法律问题

计算机取证主要是对电子证据的获取、分析、归档、保存和描述的过程，而电子证据需要在法庭上作证，因此计算机取证涉及的法律问题主要是电子证据的真实性和电子证据的证明力，其主要困难是如何证明电子证据的真实性和说明电子证据的证明力。

1. 电子证据的真实性

法律要求作为定案依据的证据应当符合真实性、合法性和关联性。电子证据若要成为法定的证据类型，关键是解决真实性的证明问题。在传统证据领域，为了保证证据的真实性，民事诉讼法和相关司法解释均要求提供证据原件即书面文件，因为原件能够保证证据的唯一性和真实性，防止被篡改或冒认。但电子证据以电磁介质为载体，没有传统观念上的原件。

对此难题的解决方法是对电子证据附加上数字签名。当证据被签发时，电子密码结合证据内容会自动生成一个新的特征码，即数字签名附加在电子证据之上，成为与证据内容不可分割的一部分。在此之后无论任何人对电子证据进行了篡改，电子证据的特征就会与原特征码不符。如果相符，则意味着电子证据未被改动。

2. 电子证据的证明力

证据的证明力指的是证据对证明案件事实所具有的效力，即该证据是否能够直接证明案件事实还是需要配合其他证据综合认定。严格来讲，“电子证据”并不是我国法律体系正式的法律用语。然而，我国相关法律法规中已经出现了与电子证据相关的内容。于 2006 年 4 月 1 日起生效的《中华人民共和国电子签名法》是我国首部对数据电文有确切描述的法律，它是一部针对电子商务发展的立法。近年来随着我国信息化的发展，不断涌现出和电子证据有关的法律案件。例如人们普遍使用手机短信进行相互联系，对于手机短信能否作为证据产生了很多争议，一些法院已经根据电子签名法认定了手机短信可以作为电子证据，并

且做出了相关判决。但是电子签名法毕竟不是专门的证据立法，其对证据制度的作用有限，电子签名法不能作为证据法的替代。1999年《中华人民共和国合同法》第11条规定："书面形式是指合同书、信件和数据电文(包括电报、电传、传真、电子数据交换和电子邮件)等可以有形地表现所载内容的形式"。该法条的规定，前者承认了电子合同的合法性，肯定了在我国民商法体系中电子证据满足书面形式的要求。《合同法》第33条还规定当事人采用数据电文订立合同可以"要求签订确认书"。另外，《合同法》的第16条、第26条、第34条规定了电子合同要约的生效时间、承诺的生效时间及合同成立地点。这些条文都涉及电子合同生效的要件，可以说是对电子合同效力的一种探索。但是这样的规定也只能是局限在民商事领域，承认了法律列举的电子证据形式可以适用书证的效力。还有某些司法解释，例如2002年4月1日生效的《最高人民法院关于民事诉讼证据的若干规定》第22条规定："调查人员调查收集计算机数据或者录音、录像等视听资料的，应当要求被调查人提供有关资料的原始载体。"从中可以看出，在民事诉讼中，最高人民法院是把视听资料做了扩大化解释，把电子证据涵盖其中，以解决司法实践中出现的立法空白。

3. 取证工具的法律效力

计算机取证常用工具有ENCASE、X-WAYS、FTK、效率源DATACOMPASS等工具。

为了确认电子证据的法律效力，还必须保证取证工具能受到法庭认可。在评估一个计算机取证的程序时，通常以Daubert测试为指导方针，主要包括4个方面：

(1) 测试。是否能够且已经测试了该程序。

(2) 错误率。程序的错误率是否已知。

(3) 公开性。程序是否已经公开并接受同等部门的评议。

(4) 可接受性。程序是否被相关的科学团体广泛接受。

随着与计算机相关的知识产权问题、不履行安全规范问题和经济诈骗等问题的增加，计算机取证显得越来越重要。法庭可命令对电子证据进行查封和分析，并调查可能是犯罪或攻击手段和工具的计算机，或是包含与刑事或民事纠纷有关的电子证据的计算机。利用专业的数据存储和恢复工具，建立相关策略，把计算机取证和入侵检测系统相结合，都是行之有效的计算机取证方法。另外，取证技术必须进一步标准化才能逐步走向成熟。为了便于其标准化，取证工具使用的程序应该被公开、被复查和被讨论。取证工具技术的公开也有助于提高工具本身的质量和实用性。

思 考 题

(1) 简述网络的类型及其特点。

(2) 导致网络脆弱的因素有哪些?

(3) 网络入侵的常用方法有哪些?

(4) 简述入侵检测系统的工作流程。

(5) 常用的入侵检测技术有哪些?

(6) 常用的计算机取证方法有哪些?

(7) 简述计算机取证的一般步骤和取证模型。

参 考 文 献

[1] 罗森林.信息系统安全与对抗技术.北京：北京理工大学出版社，2005.
[2] 蒋理，蒋真.计算机信息及网络安全实用教程.北京：中国水利水电出版社，2009.
[3] (美)William Stallings，Lawrie Brown 著.计算机安全原理与实践.贾春福，刘春波，高敏芬译.北京：机械工业出版社，2008.
[4] (美)Charles P Pfleeger，Shari Lawrence Pfleeger 著.信息安全原理与应用.李毅超，蔡洪斌，谭浩译.北京：电子工业出版社，2004.
[5] 韩最蛟，李伟.网络维护与安全技术教程与实训.北京：北京大学出版社，2006.
[6] 王玲，钱华林.计算机取证技术及其发展.软件学报，2003.
[7] Oseles L. Computer Forensics：The key to solving the crime. 2001.
[8] Sommer P. Computer Forensics：An introduction. In：Proceedings of the Compsec'92-the 9th World Conferece on Computer Security Audit and Control，1992.
[9] 陈龙，黄传河.计算机取证技术.武汉：武汉大学出版社，2007.
[10] 麦永浩，戴士剑，许榕生.计算机取证与司法鉴定.北京：清华大学出版社，2009.
[11] 陈波，于泠，肖军模.计算机系统安全原理.北京：机械工业出版社，2009.